PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 27 February 15, 1941 Number 2

Copyright 1941 by the National Academy of Sciences

NOTES ON BINARY STARS. I. ON THE SURFACE GRAVITY OF A ROTATIONALLY AND TIDALLY DISTORTED STAR

By Theodore Eugene Sterne

Harvard College Observatory

Communicated January 8, 1941

Introduction.—Tidal forces, and the rotations, cause the fluid components of a binary star to deviate from a spherical shape. Over the surfaces of such distorted stars, gravity varies. The shape and surface gravity of each component depend upon its size, the rate of rotation of the matter within it and its distribution of density, as well as upon the separation and masses. Chandrasekhar[1, 2, 3, 4] has given formulae by which both the shape and surface gravity can be computed when the distribution of density is polytropic. Elsewhere,[5] the present author has given formulae by which the shape may be computed corresponding to any distribution of density, and these formulae agree exactly with Chandrasekhar's, as they should, when the distribution is polytropic. In the present note, formulae are given for the surface gravity corresponding to any distribution of density. The surface gravity is of interest because the bolometric surface brightness is believed to be proportional to it.[6]

We use the notation and procedure of reference 5, in which formulae are given (on page 460) involving the second, third and fourth order harmonic distortions, for the shape of a rotationally and tidally distorted fluid star having any distribution of density. The method there followed employs a type of analysis first developed by Laplace. Denote the masses of the stars by m_1 and m_2; their centers of mass by Q_1 and Q_2; the distance Q_1Q_2 by R (which, in an elliptical orbit, is a function of the time). The total potential at a point A near star 1 is

$$\Psi = U + V_d \qquad (1)$$

where U is the gravitational potential of the distorted star 1, and where V_d is the disturbing potential:

$$V_d = V_t + V_r. \tag{2}$$

Here

$$V_t = \frac{m_2 G}{R} \sum_2^\infty \left(\frac{r}{R}\right)^n P_n(\cos\theta) \tag{3}$$

is the tidal potential due to star 2, and V_r is a rotational term

$$V_r = -\frac{1}{3}\omega_1^2 r^2 P_2(\cos\theta') \tag{4}$$

arising from the rotation of the matter in star 1 with a uniform angular velocity ω_1 about an axis normal to the plane of the orbit. In these expressions r is the distance Q_1A, G is Newton's constant of gravitation, P_n is the Legendre coefficient of order n, θ is the angle AQ_1Q_2, and θ' is the co-latitude. It is shown in reference 5 that if one ignores small quantities of the order of the squares and cross products of the Y's, later to be defined, then corresponding to each term of the form

$$cr^n P_n \tag{B}$$

in the disturbing potential, where $n \geqslant 2$, there is a contribution to the superficial distortion given by

$$Y_n = c\,\frac{2n+1}{n+\eta_n(a)}\,\frac{a^{n+1}}{m_1 G}\,P_n \tag{5}$$

and a contribution to the external gravitational potential of star 1 given by

$$U_n = c\,\frac{a^{2n+1}}{r^{n+1}}\,\frac{n+1-\eta_n(a)}{n+\eta_n(a)}\,P_n. \tag{6}$$

Here a is the mean radius of star 1, and the external shape of star 1 is given by

$$r = a(1 + \sum Y_n). \tag{7}$$

Different P_n's need not be referred to the same axes; and the Y_n and U_n always have the character of the surface harmonic, P_n, present in the disturbing term (B) that causes them.

In these formulae $\eta_n(a)$ is a number that depends upon n and upon the distribution of density in star 1; that is to say, upon the stellar model. It may be computed, for any particular stellar model, by solving the first order differential equation

$$r\frac{d\eta_n}{d_r} + \eta_n^2 - \eta_n - n(n+1) + \frac{6\rho}{\rho_m}(\eta_n + 1) = 0 \tag{8}$$

subject to the boundary condition that at $r = 0$, $\eta_n = n - 2$; and then $\eta_n(a)$ is the value of η_n at $r = a$. To define ρ and ρ_m precisely it is really necessary to consider the surfaces of constant density inside the distorted star. A surface of any given density, ρ, is characterized by a parameter r_1, its mean radius. In the differential equation ρ is the density at the distance r from the center, and ρ_m is the mean density interior to r, in a spherical distribution of matter, not necessarily in equilibrium, in which the density is the same function of r that ρ, in the actual distorted star, is of r_1. The specification of the stellar model may be regarded in practice as defining the ρ in equation (8), and thus ρ_m, as a function of r. For any particular model, equation (8) can always be solved easily by numerical integration so as to yield the number $\eta_n(a)$; for certain models, analytical solutions can be found.

It follows that the rotational distortion of star 1 is given by

$$\frac{r - a}{a} = Y_{\text{rot}} \tag{9}$$

and the tidal distortion by

$$\frac{r - a}{a} = Y_2 + Y_3 + Y_4 \tag{10}$$

where

$$Y_{\text{rot}} = -\Delta_2 \frac{\omega_1^2 a^3}{3m_1 G} P_2(\cos\theta'); \quad Y_n = \frac{m_2}{m_1} \Delta_n z^{n+1} P_n(\cos\theta); \tag{11}$$

and where

$$z = (a/R); \quad \Delta_n = \frac{2n + 1}{n + \eta_n(a)}; \quad P_2(x) = \frac{1}{2}(3x^2 - 1);$$

$$P_3(x) = \frac{1}{2}(5x^3 - 3x); \quad P_4(x) = \frac{1}{8}(35x^4 - 30x^2 + 3). \tag{12}$$

Values of Δ_2, Δ_3 and Δ_4 are given by Chandrasekhar[2, 4] for the polytropes; he obtained the values not by solving the present equation (8) but by special analytical procedures. It has been mentioned that the squares and cross products of the Y's have been neglected in the present treatment. The smallest square or cross-product term to be neglected is the square of a Y_2 term, and this square is of the same order of magnitude as Y_5. For most purposes the present degree of accuracy, which is the same as Chandrasekhar's for polytropes, is sufficient. If a still higher degree of accuracy were desired, it would be necessary to take into account not only the squares and cross products, and harmonic distortions with orders higher than the fourth, but also corrections to V_i arising from the distortion of

star 2 produced by its own rotation and the tidal action of star 1. In particular, there is little to be gained by considering the square of a Y_2 term without considering Y_5, and little to be gained by considering Y_5 without taking into account the squares of the Y_2 terms. To the order of accuracy of the present treatment, the various Y's and U's are additive.

Consider, now, a term in the disturbing potential

$$cr^n P_n(\cos\theta) \qquad (B)$$

where θ is measured from any appropriate axis—in the case of a tidal term it must be measured from the line of centers, and in the case of a rotational term from the axis of rotation. Let us consider the effect of this term upon the surface gravity. Considering only this single disturbing term we shall have for the total potential in the external vicinity of star 1

$$\Psi = \frac{m_1 G}{r} + U_n + cr^n P_n(\cos\theta)$$

in which the first term on the right is the potential due to the undisturbed star 1. The corresponding value of apparent gravity, g, at the surface of star 1 is of course

$$g = -\frac{d\Psi}{dn}$$

evaluated at the surface, where dn is measured in the direction of the outward normal. The value of gravity being continuous at the surface, we do not need the value of Ψ at interior points; and g is the resultant of the disturbing force and the attraction of the distorted star 1. The outward normal makes a small angle with the radius vector r. It is easy to show that this angle is of the order of magnitude of Y_2 or smaller, so that its cosine can differ at most from unity by a quantity of the order of magnitude of the square of Y_2. To the present accuracy, therefore,

$$g = -\frac{d\Psi}{dr}$$

evaluated at $r = a(1 + Y_n)$. Making use of the expressions already given for U_n and Y_n, one finds after some reduction that

$$\frac{g - g_0}{g_0} = -[1 + \eta_n(a)]\, Y_n \qquad (13)$$

where

$$g_0 = \frac{m_1 G}{a^2} \qquad (14)$$

and is a constant.

More generally, if a number of terms of the form (B) in the disturbing function are to be considered, then

$$\frac{g - g_0}{g_0} = -\sum[1 + \eta_n(a)]\, Y_n. \tag{15}$$

This equation has a simple meaning: for each separate harmonic distortion there is a corresponding variation of surface gravity, such that the fractional variation of surface gravity is proportional over the star's surface to the fractional variation in the radius vector. The factor of proportionality is $1 + \eta_n(a)$ and depends on the stellar model and on the order of the distortion.

A Theorem.—From equations (5) and (13), and from the form of the term (B), it follows that the apparent surface gravity, corresponding to any harmonic distortion of order n, obeys the law

$$\frac{g - g_0}{g_0} = -\left(2 + \frac{1}{n}\right)\varphi_n + (n - 1)\, Y_n, \tag{15'}$$

where φ_n is the ratio of the outward normal component of the disturbing force at the surface to g_0. Equation (15′), not used here, enables one to compute the value of apparent gravity over the surface of any tidally and rotationally distorted fluid mass from a knowledge of its shape and of the values of the disturbing forces at the surface, even when the internal constitution is unknown. If the actual surface distortion is analyzed into zonal (or tesseral) harmonics of different orders, Y_n, then the total $(g - g_0)/g_0$ is merely the sum of the right-hand member of (15′) over all the zonal (or tesseral) harmonics that are present. When n is set equal to two, equation (15′) becomes Clairaut's well-known relation that is used in finding the earth's ellipticity from measures of gravity. I am indebted to Dr. Russell for having pointed out to me that some general relation—such as the exceedingly general equation (15′)—was necessarily contained in the analysis of this Note and in that of reference 5.

Surface Gravity on a Component of a Binary Star.—In the practical application of equation (15), there are two Y's of the second order, namely Y_{rot} and the tidal Y_2. There are also the tidal Y_3 and Y_4 of the third and fourth orders, respectively. Thus, in a binary star, the surface gravity g of component 1 is given by

$$\frac{g - g_0}{g_0} = -[1 + \eta_2(a)]\, Y_{\text{rot}} - \sum_2^4 [1 + \eta_n(a)]\, Y_n \tag{16}$$

where the Y's are given by equation (11).

The $\eta_n(a)$'s are related to the k_n's, of importance in the theory of apsidal motion, in the manner indicated in reference 5. For all models,

$$k_n = \frac{n + 1 - \eta_n(a)}{2n + 2\eta_n(a)} \text{ and } \Delta_n = 1 + 2k_n.$$

The "K" of Russell's paper[6] is the present k_2; Russell does not consider the third and fourth order harmonic distortions. The quantity

$$\Delta_n[1 + \eta_n(a)] = n + 2 - 2k_n(n - 1)$$

gives the dependence of the range of variation of surface gravity, over the surface of a star, upon the distribution of density and upon the order of the harmonic distortion. The $\eta_n(a)$'s can be found for any density distribution by solving equation (8), as has been mentioned; and they can be found without further work for polytropes by using Chandrasekhar's Δ's and the relations (12). For the reader's convenience we list in the accompanying table the values of k, Δ, $1 + \eta(a)$ and $\Delta[1 + \eta(a)]$ for polytropes. The polytrope of index zero is a homogeneous model; that of index five is a completely concentrated, Roche, model. The values for any model, polytropic or not, should lie between the values corresponding to the polytropic indices zero and five, since these are limiting distributions of density. Observational evidence, obtained principally by Cowling,[7] Sterne[5, 8] and

POLYTROPIC INDEX	0	1	1.5	2	3	4	5
k_2	$^3/_4$	0.2599	0.1446	0.0741	0.0144	0.00134	0
Δ_2	$^5/_2$	1.5198	1.2892	1.1482	1.0289	1.00267	1
$1 + \eta_2(a)$	1	2.2898	2.8784	3.3546	3.8596	3.9867	4
$\Delta_2[1 + \eta_2(a)]$	$^5/_2$	3.4802	3.7108	3.8518	3.9711	3.9973	4
k_3	$^3/_8$	0.1064	0.0540	0.0244	0.00368	0.00024	0
Δ_3	$^7/_4$	1.2129	1.1079	1.0488	1.00736	1.00047	1
$1 + \eta_3(a)$	2	3.7713	4.3183	4.6743	4.9489	4.9967	5
$\Delta_3[1+\eta_3(a)]$	$^7/_2$	4.5742	4.7842	4.9024	4.9853	4.9991	5
k_4	$^1/_4$	0.0602	0.0281	0.0116	0.00140	0.00007	0
Δ_4	$^3/_2$	1.1205	1.0562	1.0231	1.00281	1.00014	1
$1 +\eta_4(a)$	3	5.0321	5.5211	5.7968	5.9748	5.9987	6
$\Delta_4[1 + \eta_4(a)]$	$^9/_2$	5.6385	5.8314	5.9307	5.9916	5.9996	6

Russell[6] from the apsidal motions, suggests that actual stars are considerably concentrated. They appear to be sufficiently concentrated for values of Δ, $1 + \eta(a)$ and $\Delta[1 + \eta(a)]$ to be taken from the last, Roche, column of the table with sufficient accuracy for most practical purposes. In many practical applications, moreover, it will be legitimate, in the absence of conflicting observational evidence, to set the constant angular velocities ω_1 and ω_2 equal to the mean orbital motion $2\pi/P$ where P is the period.

Summary.—In *Monthly Notices*, **99**, 460 (1939), the author gave formulae for predicting the shape of a rotationally and tidally distorted fluid component of a binary star. The component could have any distribution

of density, and harmonic distortions through the fourth order were taken into account. In the present note formulae, valid for any distribution of density, are given for the surface gravity. A tabular compilation is made of the numerical coefficients that are needed to compute the apsidal motions, distortions and surface gravities of some polytropic models.

[1] Chandrasekhar, *M. N.*, **93**, 390 (1933).
[2] Chandrasekhar, *Ibid.*, **93**, 449 (1933).
[3] Chandrasekhar, *Ibid.*, **93**, 462 (1933).
[4] Chandrasekhar, *Ibid.*, **93**, 539 (1933).
[5] Sterne, *Ibid.*, **99**, 451 (1939).
[6] For references see Russell, *Ap. J.*, **90**, 641 (1939).
[7] Cowling, *M. N.*, **98**, 734 (1938).
[8] Sterne, *Ibid.*, **99**, 662 (1939).

NOTES ON BINARY STARS. II. ON THE LIGHT OF ECLIPSING BINARIES OUT OF ECLIPSE

BY THEODORE EUGENE STERNE

HARVARD COLLEGE OBSERVATORY

Communicated January 8, 1941

Introduction and the Second Order Harmonic.—In Note I in these PROCEEDINGS,[1] formulae were given for finding the surface gravity of a rotationally and tidally distorted component, of a binary star, having any distribution of density. Here we give a brief derivation of the effects of the distortion, and of the variation over the surface of the surface brightness, upon the observable light of the components outside of eclipse. The method used is simple and general, and will have applications to other phenomena involving the light of a distorted star having a non-uniform surface brightness.

A theorem was proved by von Zeipel[2] and Chandrasekhar[3] that over a tidally and rotationally distorted star, the bolometric surface brightness must be proportional to gravity if the star is in hydrostatic equilibrium. The argument can be put in a simpler and more physical form than the proofs given by the two previously mentioned authors. The equation of hydrostatic equilibrium is $dP = \rho d\Psi$, where P is the pressure, ρ is the density and Ψ is the total potential as in Note I. It follows that P and ρ must be functions of Ψ alone in a star that is in hydrostatic equilibrium under the influence of its own gravity and of any disturbing potential. Hence the temperature, T, must also be a function of Ψ alone and so must be the opacity since that is a function of T and ρ. The outward flow of

heat is given by the product of the temperature gradient and a function of the temperature, density and opacity. The function is a constant over any equipotential surface in the star, and thus the outward flow of heat across such a surface is proportional, over the surface, to the temperature gradient. Since the equipotential surfaces are isothermal surfaces, the temperature gradient is proportional over such a surface to the gradient of Ψ and the latter gradient is merely the value of apparent gravity, g. Thus the outward flow of heat is proportional, over any equipotential surface within the star, to gravity; and if one takes for the selected equipotential surface the outer surface of the star, or a surface close within it, it follows that the bolometric surface brightness is proportional to apparent surface gravity.

The preceding physical demonstration reveals the limitations of the theorem. The theorem need not hold if there are convection currents near the outer surface, for then the equation of hydrostatic equilibrium is inapplicable. The equation will be in error by amounts depending upon the acceleration of the stellar matter near the selected equipotential surface. It does not seem likely that this source of error can be very serious; a more serious one is that an appreciable transport of heat by material currents may render the flow of heat no longer proportional, over the selected equipotential surface, to the temperature gradient. It is difficult to predict theoretically, for actual stars, how much the bolometric surface brightness deviates from the strict proportionality, to surface gravity, that would hold under conditions of perfect hydrostatic equilibrium. Moreover, even if exact proportionality holds for the bolometric surface brightness, it can hardly hold for the surface brightness in a single wave-length, for the quality of the radiation must vary over the distorted surface, with the surface brightness.

Let us take for the variation of surface brightness, as it is influenced by the presence of a harmonic distortion of order n, the expression

$$J = J_0(1 - \beta_n Y_n).$$

Here J is the surface brightness as influenced by this distortion, J_0 is the average surface brightness and Y_n (defined in Note I) is such that the radius r is given by

$$r = r_0(1 + Y_n) \tag{17}$$

where r_0 is the average radius. If the surface brightness is bolometric, and if the theorem is applicable, then the coefficient β_n is exactly equal to the quantity $1 + \eta_n(a)$ whose value has been found, as a function of the stellar model and of the order n of the harmonic distortion, in Note I. To take account of possible deviations from the theorem, the safest course appears to be to set

$$\beta_n = y[1+\eta_n(a)] \tag{18}$$

where y is the coefficient by which the full variation, predicted by the theorem, must be multiplied to obtain the actual variation of surface brightness. If the theorem is applicable, $y = 1$; if the surface brightness is independent of surface gravity, $y = 0$. The coefficient y is the same as Russell's,[4] and some observational evidence that the bolometric surface brightness follows the law $y = 1$ was presented by Zdeněk Kopal to the New York Academy of Sciences in a lecture given on October 20, 1939, and is contained in a paper, now in press, that he has been kind enough to show me.

We also take into account the darkening of an apparent stellar disc towards its limb by using the usual law of limb-darkening that renders the apparent surface brightness proportional to $1 - x + x \cos \gamma$, where γ is the angle between the outward normal and the line of sight. Thus

$$J = J_o(1 - \beta_n Y_n)(1 - x + x \cos \gamma), \tag{19}$$

where J_o is the average normal surface brightness, for each harmonic component of the distortion separately.

As in Note I, we shall ignore the squares and cross products of the Y's, and to this order of accuracy the fractional effects of the different distortions are additive. As pointed out previously,[1] it is useless to consider small quantities of the order of the square of the oblateness or higher, unless harmonic distortions of orders higher than four are considered; here we shall consider harmonic distortions of orders up to and including the fourth. To the present accuracy, no squares or cross products of the Y's can appear in any results. The rotational distortion Y_{rot} hardly concerns us, for it can introduce no observable effects in the light out of eclipse. Out of eclipse, the aspect of the rotational ellipsoid does not change since the direction of the axis of rotation does not change. Thus the observed variation in light does not involve any effect proportional purely to Y_{rot}. Moreover, Y_{rot} can combine with another distortion Y_n only in the form of a product $Y_{\text{rot}} Y_n$, and we have agreed to ignore such small cross-product terms. Since we are not concerned with the light during eclipse, we can thus legitimately ignore the rotational distortion so far as changes of light are concerned. By similar reasoning we see that we can consider the effect of each tidal distortion Y_n separately, as though it alone were present, and we proceed first to consider Y_2, the most important tidal distortion.

The tidal distortion Y_2 causes the surface of star 1 to become approximately a prolate ellipsoid, with its long axis along the line of centers.* The radius to any point on the surface is given by

$$Y_2 = (2/3)\epsilon P_2(\cos \delta')$$

so

$$r = 1 + (2/3)\epsilon P_2(\cos \delta') \tag{20}$$

correctly to the first order in the tidal oblateness, ϵ, if the mean radius is taken as unity. Here δ' and ψ are "astrocentric" spherical polar coördinates in which δ' is the angle between r and the line of centers. The polar coördinates of the outward normal to the surface at the given point may be denoted by δ and ψ; and then in terms of the "astrographic" coördinates δ and ψ it is easy to find expressions for the radius vector, r, and the element of surface, dS. Since

$$\frac{dr}{rd\delta'} = -2\epsilon \sin \delta' \cos \delta' \tag{21}$$

one has to the first order in ϵ

$$\delta' = \delta - 2\epsilon \sin \delta \cos \delta \tag{22}$$

and so $r = 1 + (2/3)\epsilon P_2(\cos \delta)$. Moreover, since the outward normal makes an angle to the radius vector that is proportional to ϵ, the cosine of the angle is unity to the order of ϵ. Thus

$$\begin{aligned} dS &= r^2 \sin \delta' d\delta' d\psi \\ &= [1 - (8/3)\epsilon P_2(\cos \delta)] \sin \delta d\delta d\psi \end{aligned}$$

or

$$dS = [1 - (8/3)\epsilon P_2(\cos \delta)] d\Omega \tag{23}$$

correctly through terms of order ϵ. Here $d\Omega$ is an element of solid angle in the normal system δ, ψ.

Take a right-handed system of axes X, Y, Z such that the Z-axis is the axis of the orbit and the Y-axis is in the direction of the ascending node of the relative orbit of star 1. Let the inclination be denoted as usual by i, and let θ be the true orbit longitude of star 1 measured from superior conjunction. When the orbit is circular, this θ is the usual θ of eclipsing star theory.[8] Take also a set of rotating axes X', Y', Z' with Z' towards the earth, and X' in the plane of the sky and towards the apparent position of star 2. The angle α between the Z'-axis and the line of centers is given by

$$\cos \alpha = \sin i \cos \theta. \tag{24}$$

We employ a set of spherical polar angles γ, φ in the primed system, such that γ is the polar angle measured from the Z'-axis. In the primed system, the direction cosines of the direction γ, φ are $\sin \gamma \cos \varphi$, $\sin \gamma \sin \varphi$, $\cos \gamma$; those of the line of centers are $\sin \alpha$, 0, $\cos \alpha$. Hence

$$\cos \delta = \sin \alpha \sin \gamma \cos \varphi + \cos \alpha \cos \gamma. \tag{25}$$

The angle γ is just the angle involved in the limb-darkening; and $d\Omega = \sin\gamma d\gamma d\varphi$. Accordingly, the light received at the earth from the element of surface corresponding to $d\gamma$ and $d\varphi$ on star 1 is proportional to

$$dl = J_o[1 - x + x\cos\gamma][1 - (2/3)\epsilon(4 + \beta_2)P_2(\cos\delta)] \cdot \sin\gamma\cos\gamma\, d\gamma d\varphi \quad (26)$$

and the total light is proportional to this quantity integrated over all values of γ from 0 to $\pi/2$ and of φ from 0 to 2π. The integration is easily effected since $\cos\delta$ is given by (25); the result is that the light is proportional to

$$\pi J_o\left[1 - \frac{x}{3} - \frac{2}{3}\epsilon(\beta_2 + 4)P_2(\cos\alpha)\left(\frac{1-x}{4} + \frac{4}{15}x\right)\right]$$

so that one can write for the light of star 1

$$L = L_o\left[1 - \epsilon\frac{(\beta_2 + 4)(15 + x)}{30(3 - x)}P_2(\cos\alpha)\right] \quad (27)$$

correctly to terms of order ϵ where L_o is a constant, and if we are concerned only with the changes in light caused by the changing geometry† during an orbital period we can write this as

$$L = L_o'\left[1 - \epsilon\frac{(\beta_2 + 4)(15 + x)}{20(3 - x)}\sin^2 i\cos^2\theta\right] \quad (27')$$

where L_o' is a new constant. Equation (27′) is equivalent to Russell's equation (14) of reference 4, since his term that involves the rotational ellipticity is a mere constant whose effect can be absorbed, to his accuracy and ours, in the constant L_o'.

If "gravity darkening" is ignored and there is no darkening towards the limb, then β_2 and x are both zero and (27′) agrees to the order of ϵ with Russell's equation (41) of reference 9 in which his "ϵ" is an eccentricity and his "ϵ^2" is our 2ϵ. If gravity darkening is ignored and there is complete darkening towards the limb then β_2 is zero and x is unity and equation (27′) agrees as it should with Russell and Shapley's equation in the text on page 400 of reference 8. In accordance with the definition of β_2 and the results in the previous Note, the full gravity effect ($y = 1$) corresponds to a value of β_2 equal to four for the Roche model and to a value (which in any particular case can be exactly computed as in Note I) of nearly four for any strongly concentrated model. As noticed by Russell,[4] therefore, a full gravity effect must in practice nearly double the variation in light, observed outside of eclipse, that arises from the principal tidal distortion Y_2.

An equation similar to (27) holds, of course, for star 2; and the observed light outside of eclipse is the combined light.

The Third and Fourth Order Harmonics.—For the third order harmonic, one has

$$r = \epsilon_3 P_3(\cos \delta') \tag{28}$$

where ϵ_3 is a numerical coefficient that may be found as described in Note I. By the same procedure as before, one finds that

$$dS = [1 - 10\epsilon_3 P_3(\cos \delta)]d\Omega \tag{29}$$

and thus the light of star 1 outside of eclipse, as influenced by the third order tidal deformation, is given by

$$\int_0^{2\pi}\int_0^{\pi/2} [1 - x + x \cos \gamma][1 - \epsilon_3(10 + \beta_3)P_3(\cos \delta)] \sin \gamma \cos \gamma d\gamma d\varphi$$

and thus the light of component 1 varies in accordance with the equation

$$L = L_o\left[1 - \epsilon_3 \frac{(10 + \beta_3)x}{4(3 - x)} P_3(\sin i \cos \theta)\right]. \tag{30}$$

Here $\beta_3 = y[1 + \eta_3(a)]$.

For the fourth order harmonic tidal deformation,

$$r = 1 + \epsilon_4 P_4(\cos \delta') \tag{31}$$

where ϵ_4 is a numerical coefficient given in Note I, and one readily finds that

$$dS = [1 - 18\epsilon_4 P_4(\cos \delta)]d\Omega. \tag{32}$$

Thus the light of star 1 outside of eclipse, as influenced by the fourth order tidal distortion, is proportional to

$$\int_0^{2\pi}\int_0^{\pi/2} [1 - x + x \cos \gamma][1 - \epsilon_4(18 + \beta_4)P_4(\cos \delta)] \sin \gamma \cos \gamma d\gamma d\varphi$$

so that the light varies in accordance with the equation

$$L = L_o[1 + \frac{\epsilon_4(18 + \beta_4)(1 - x)}{8(3 - x)} P_4(\sin i \cos \theta)]. \tag{33}$$

We notice that no variation of light arises from the third harmonic unless there is some limb-darkening, nor any from the fourth unless there is incomplete limb-darkening. These particular properties of the odd and even harmonics higher than the second were noticed by Russell in a paper[10] on the light-variation of asteroids.

If there is a complete gravity effect, then from Note I we see that β_3 and β_4 are nearly five and six, respectively, for strongly concentrated stars. The exact values can be computed for any particular stellar model. The full effect of the variation of gravity is thus in general to multiply the varia-

tions of light outside of eclipse, arising from the third and fourth order distortions, by factors of 3/2 and 4/3, respectively.

If the right-hand members of equations (27), (30) and (33) are written as $L_o(1 + \lambda_2)$, $L_o(1 + \lambda_3)$ and $L_o(1 + \lambda_4)$, then the combined variations give for the light of star 1 the quantity $L_o(1 + \lambda_2 + \lambda_3 + \lambda_4)$. A similar result holds for star 2.

One must remember, when combining the lights of stars 1 and 2, that the "θ" for star 2 differs by 180° from the θ for star 1. Thus the second and fourth order variations of star 2 are given by equations (27) and (33) with appropriate coefficients, but in the third order variation of star 2, given by equation (30) with appropriate coefficients, P_3 must be replaced by $-P_3$ if for θ one employs the θ of this Note, which is the θ of star 1. If the components are identical, one sees that no odd-order harmonics can appear in the combined light whatever may be the law of darkening; for the effects of the odd harmonics of star 1 are then precisely neutralized in the combined light by those of star 2. On the other hand, the even-order harmonics of stars 1 and 2 must always reinforce each other.

The effects of reflection must of course be added to the effects considered here, in order to obtain a complete and accurate representation of the light outside of eclipse.

Systematic Effects.—The quantity L_o is the light of star 1 free from the effects of the tidal distortions. Since the integral of a surface harmonic over a sphere is zero if the order differs from zero, L_o is thus dependent merely upon the luminosity of star 1 and its distance from the earth—except for reflection and the small effect considered shortly. The quantity L_o' of equation (27′) contains, however, a systematic tidal effect arising from the constant term in $P_2(\cos \alpha)$; and it should not be used without correction to obtain a luminosity free from tidal effects. It can easily by corrected by the equation

$$L_o = L_o'\left[1 - \epsilon\frac{(\beta_2 + 4)(15 + x)}{60(3 - x)}\right]. \tag{34}$$

The rotational distortion does not influence the variation in light outside of eclipse to the present order of accuracy, but it tends to introduce a systematic error in the luminosity. This error can be allowed for by the relation (obtained by the same general procedure as equation (27))

$$L_o = L^*\left[1 + \epsilon_{\text{rot}}\frac{(\beta_2 + 4)(15 + x)}{30(3 - x)}P_2(\cos i)\right] \tag{35}$$

where L^* is free from the effects of the unequal distribution of brightness over the surface of the rotational ellipsoid. Here ϵ_{rot} is the rotational

oblateness, equal to $(-3/2)$ times the coefficient of P_2 in Y_{rot}, given in equation (11) of Note I.

Summary.—Equations are obtained for the light, as influenced by rotational and tidal distortions, of a component of an eclipsing star outside of eclipse. The results are valid for any distribution of density, and all harmonic distortions are taken into account with orders up to and including the fourth. The conditions of applicability of von Zeipel's theorem about surface brightness are examined, and some systematic corrections are considered to the observed absolute magnitudes of rotating stars.

* In the limit where the orbital period of a binary star is long compared with the free periods of spherical harmonic oscillation of the components, the tidal deformations are given by the equilibrium theory of the tides. See Cowling, reference 5, and Sterne reference 6. For the most complete discussion of the ratio of periods see Sterne, reference 7, pages 664 and 665. In the present Notes, the equilibrium results are used in which the line of centers is a common axis of all the tidal harmonics.

† Changes in light caused by changes in e during the motion will be studied in a future Note.

[1] Sterne, *Proc. Nat. Acad. Sci.*, **27**, 93 (1941).
[2] von Zeipel, *M. N.*, **84**, 702 (1924).
[3] Chandrasekhar, *Ibid.*, **93**, 539 (1933).
[4] Russell, *Ap. Jour.*, **90**, 641 (1939).
[5] Cowling, *M. N.*, **98**, 734 (1938).
[6] Sterne, *Ibid.*, **99**, 451 (1939).
[7] Sterne, *Ibid.*, **99**, 662 (1939).
[8] Russell and Shapley, *Ap. Jour.*, **36**, 404 (1912).
[9] Russell, *Ibid.*, **36**, 54 (1912).
[10] Russell, *Ibid.*, **24**, 1 (1906).

NOTES ON BINARY STARS. III. THE ELLIPTICITY EFFECT IN ECCENTRIC ECLIPSING BINARIES

BY THEODORE EUGENE STERNE

HARVARD COLLEGE OBSERVATORY

Communicated January 11, 1941

In Note II we derived[2] equations for the variation in the light of a binary star outside of eclipse, as influenced by the distortion and by the nonuniform distribution of surface brightness. The tidal distortion and the distribution of surface brightness depend in part, as has been shown in Notes I[1] and II, upon the separation. When we computed the light in Note II, we considered the distortion and surface brightness to stay constant during the motion, as they would in a circular orbit. If the orbit is eccentric, however, the separation changes and causes the distortion and

distribution of surface brightness to change, in the period of the orbit. Here we give the simple equations for the effect of these changes upon the light outside of eclipse.*

The light of star 1, as influenced by the harmonic tidal distortion Y_2 of order two, is given outside of eclipse by equation (27) of Note II. In that equation, ϵ is the tidal oblateness of the second order; it is three-halves of the coefficient of P_2 in the expression (11) for Y_2 in Note I. Let v denote the true anomaly, $\overline{\omega}$ the longitude of periastron, and e the eccentricity of the relative orbit, of star 1. Then if the mean distance between centers of mass is A, $(A/R) = (1 + e\cos v)/(1 - e^2)$. But by equation (11) of Note I, ϵ varies as $(1/R)^3$; thus $\epsilon = \epsilon'(A/R)^3$ where ϵ' is the value of ϵ at the instants of mean separation and is given by

$$\epsilon' = (3/2)(m_2/m_1)\,\Delta_2(a/A)^3. \tag{36}$$

Thus by equation (27) of Note II one has for the light of star 1

$$L = L_0\left[1 - \epsilon'\frac{(\beta_2 + 4)(15 + x)}{60(3 - x)}(3\sin^2 i\cos^2\theta - 1)\frac{(1 + e\cos v)^3}{(1 - e^2)^3}\right] \tag{37}$$

where

$$\theta = \overline{\omega} + v - 90°$$

and where L_0 is the light of star 1 free from the effects of the second order tidal distortion.

In equation (37), ϵ' is a constant during the motion; and so is the whole coefficient $\epsilon'(\beta_2 + 4)(15 + x)/60(3 - x)$, whose value may be denoted by $S_{2,1}$. The combined lights of stars 1 and 2 outside of eclipse, as influenced by the changing tidal distortion of the second order, will thus obey the relation

$$L_{\text{combined}} = 1 - S_2(3\sin^2 i\cos^2\theta - 1)\frac{(1 + e\cos v)^3}{(1 - e^2)^3} \tag{38}$$

where S_2 is a mean of $S_{2,1}$ and $S_{2,2}$ (for star 2), weighted with the L_0's of the two stars, and should be observationally determinable. In (38) we have taken the combined L_0's as unity.

In the same way, the third and fourth order harmonic tidal distortions produce light variations for star 1 of

$$L = L_0\left[1 - S_{3,1}(5\sin^3 i\cos^3\theta - 3\sin i\cos\theta)\frac{(1 + e\cos v)^4}{(1 - e^2)^4}\right] \tag{39}$$

and

$$L = L_0\left[1 + S_{4,1}(35\sin^4 i\cos^4\theta - 30\sin^2 i\cos^2\theta + 3)\frac{(1 + e\cos v)^5}{(1 - e^2)^5}\right], \tag{40}$$

respectively, outside of eclipse. Here the coefficients are given by the equations, obtained from Notes I and II,

$$S_{3,1} = \frac{m_2}{m_1}\Delta_3\left(\frac{a}{A}\right)^4\frac{(10+\beta_3)x}{8(3-x)};\quad S_{4,1} = \frac{m_2}{m_1}\Delta_4\left(\frac{a}{A}\right)^5\frac{(18+\beta_4)(1-x)}{64(3-x)} \tag{41}$$

and remain constant during the motion. Equations (39) and (40) hold for the combined lights of the two stars, if the L_0's are set equal to unity and the S's are replaced by S_3 and S_4, the weighted means, respectively, of $S_{3,1}$ and $-S_{3,2}$, and of $S_{4,1}$ and $S_{4,2}$.

In this Note, we content ourselves with giving the theory of the "distortion" terms only, in the light. To the changes in the combined light of the components due to the distortions, there must, of course, be added the changes in light due to the reflection effect, if one wishes to obtain an exact and complete representation of the observed light outside of eclipse; and in the reflection effect, the changing separation must be taken into account.

We notice that equations (38) and (40) for the combined lights both predict small differences (in the same sense) between the combined lights at the ascending and descending nodes ($\theta = 270°$ and $\theta = 90°$), unless $\bar{\omega}$ is either 90° or 270°. The reflection effect, being greater near periastron than near apastron, should always reinforce this difference, which the third tidal harmonic does not influence.

The author has been informed that Dr. Kopal has a paper in press on the subject of this Note—the effect of changes in figure upon the light outside of eclipse. Dr. Kopal's investigations, and the author's, have been independent. The present Note contains, substantially, the contents of a paper that the author read on December 30, 1940, before a meeting of the American Astronomical Society in Philadelphia.

Summary.—The effects are predicted of changes in figure, occurring during a revolution in an eccentric orbit, upon the light outside of eclipse. Ordinary and "gravitational" darkening are taken into account, and the results are applicable to any distribution of density.

* Cowling[3] has considered the effect of the changes in the second order tidal distortion, and the author[4] has considered the effect of the changes in the second, third and fourth order tidal distortions, upon the longitude of periastron. The theory of the effects of the changing distortions upon the light is, needless to say, considerably simpler than the theory of their effects upon the orbit.

[1] Sterne, *Proc. Nat. Acad. Sci.*, **27**, 93–99 (1941).
[2] Sterne, *Ibid.*, **27**, 99–106 (1941).
[3] Cowling, *M. N.*, **98**, 734 (1938).
[4] Sterne, *M. N.*, **99**, 451 (1939).

BIVALENT STRUCTURE IN THE FLY, MELOPHAGUS OVINUS L. (PUPIPARA, HIPPOBOSCIDAE)

BY KENNETH W. COOPER

DEPARTMENT OF BIOLOGY, PRINCETON UNIVERSITY

Communicated December 20, 1940

The customary absence of genetic crossing-over in male *Drosophila* made difficult the acceptance of any hypothesis of meiosis in which the chiasma is the condition *sine qua non* of association of homologous chromosomes in bivalents at the first meiotic metaphase. Darlington (1931) recognized the difficulty inherent in *Drosophila* males and suggested that meiosis of the male brachycerous Diptera is exceptional not in the absence of chiasmata, and hence of crossing-over, but in that the chiasmata present are invariably double, reciprocal,[1] and located in genetically inert regions of the chromosomes. However, this view was later retained only for the sex chromosomes, for his own work (1934) as well as that of Keuneke (1924), Metz (1926) and Bauer (1931) seemed to show that the association of homologous autosomes in higher Diptera is derived from an exaggerated expression of the forces of somatic pairing. Since physical connections have not been generally observed between the four chromatids in autosomal bivalents, Darlington holds the forces of repulsion between homologous chromosomes to be in exact equilibrium with forces of attraction at meiotic metaphase. The absence of genetic crossing over in the male brachycerous fly, therefore, is currently believed by Darlington to be due to an anomalous meiosis in which the autosomes are held together as bivalents only by an abnormally intense expression of the attractive forces of homologous genes, and the sex chromosomes by the occurrence of nearly invariably reciprocal chiasmata in the genetically inert but homologous regions.

Of the suborder Brachycera of the Diptera, especial interest lies in the section *Pupipara*.[2] Not only have the Pupipara remained unexamined cytologically, but they represent perhaps the biologically most specialized derivatives of the Schizophora—the very series to which *Drosophila* belongs. The present paper gives a preliminary account of certain critical features in the meiosis of the male of *Melophagus ovinus* L.—the sheep tick, or ked—with a brief discussion of their significance for the general theory of the meiosis of male Brachycera. It will be shown that not only does *Melophagus* possess a unique type of autosomal bivalent for flies, but that the structure of these autosomal bivalents sheds new light upon the interpretation of the structure of sex chromosome bivalents in *Drosophila*.

The gonial metaphase plate of the male *Melophagus ovinus* (Fig. 2) shows 18 chromosomes.[3] The diploid set may be resolved into a group of eight

large autosomes, eight small autosomes and an unequal XY pair. The large autosomes are all of similar size and each bears a trabant-like arm proximal to its kinetochore. The small autosomes likewise are all of similar proportions, and each bears a trabant-like region distally and, perhaps, another proximal to its kinetochore. Whereas the autosomes all show a more or less paired association on the mitotic metaphase plate, the sex chromosomes evince no such tendency (Fig. 2).

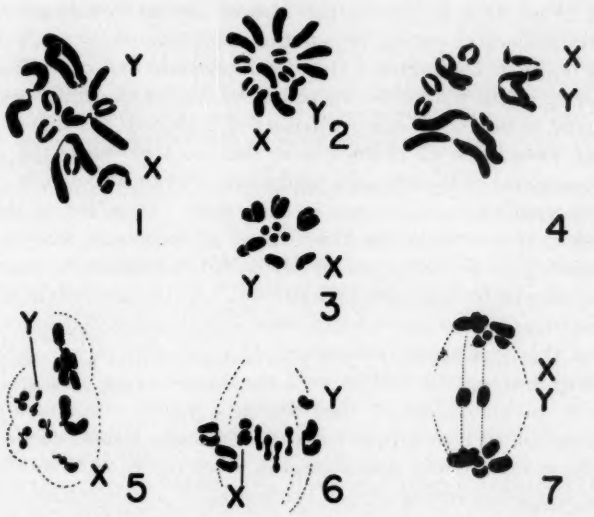
The most significant peculiarity of the somatic mitoses (including gonial and embryonic cells, and follicle cells of the ovary) is the derivation of the paired condition of metaphase from a proximal association or pairing in the large autosomes at prophase (Figs. 1, 4). The paired regions are terminal,[4] proximal to the kinetochore and may show faintly staining wisps connecting the homologues.[5] Other than by proximal terminal pairing, no tendency toward somatic pairing is shown by these prophase chromosomes. The expressions of affinity are localized in the proximal trabant-like regions, and the remainders of homologous chromosomes appear mutually indifferent.

In the small autosomes the condition is not so clear. Frequently both ends of homologous small chromosomes appear paired, while the mid-regions bow away from one another as though in a state of active repulsion. Not uncommonly small autosomes are paired at one end only; in such cases the paired ends are clearly those distal to the kinetochore. At no stage do the sex chromosomes evince any tendency to undergo somatic pairing in the male (Figs. 1, 2, 4).

In the prophase of the first spermatocyte there has been found no interpretable evidence of the early synaptic stages of meiosis. In this respect *Melophagus* resembles other cyclorrhaphous Brachycera.[6] At diakinesis, however, the situation is strikingly different from that described for other flies such as *Drosophila* which have exceedingly small chromosomes. Both classes of autosomes clearly show chromatic (feulgen-staining) connections as the mode of bivalent association. Although the large autosomes condense in a common karyosomic cluster and only infrequently show their structure clearly, the small autosome bivalents are quite evenly distributed throughout the nucleus and easily studied. At *late* diakinesis the small bivalents are held together by distal connections, whereas in the large autosomes there are similar proximal unions (Fig. 5). As the first metaphase plate forms, the repulsions of the kinetochores become especially evident and clearly locate these points of association (Fig. 6). Before the plate is fully formed the associative strands of the bivalents become attenuated (Fig. 6) and may finally appear to have disjoined. The result is a metaphase plate in which the bivalents appear to be held together only by Darlington's forces.

The sex chromosomes during all stages of meiosis remain entirely apart

from one another and show no evidences of association (Figs. 3, 5, 6). Their behavior is unique among Brachycera, but resembles that of the supernumerary chromosomes of *Tipula* (Bauer, 1931) and the sex chromosomes of *Fungivora*, *Mycetophila* and *Exechia* (Frolova, 1929), for they remain in the equator during early anaphase (Fig. 7). Only after the autosomes have attained the poles do the sex chromosomes reductionally separate.



SPERMATOGENESIS OF *Melophagus ovinus* L.; FIGURES SEMI-DIAGRAMMATIC

Figure 1, gonial prophase. Figure 2, gonial metaphase. Figure 3, polar view of first meiotic metaphase plate. Figure 4, gonial prophase. Figure 5, diakinesis; one small autosome bivalent missing. Figure 6, prometaphase of first meiotic division; chromosomes have been spread across equator for purposes of clarity. Figure 7, anaphase of first meiotic division. It must be noted that figures 1 and 4 are approximately one-third again the magnification of the remaining figures.

The second spermatocyte division is interesting in that the small autosome univalents separate first, then the sex chromosomes divide equationally and lastly the large autosomes undergo anaphase. Although there is thus a definite order in which each chromosome class separates, the chromosomes all reach the poles simultaneously.

The chief importance of the striking chromosome behavior of male

Melophagus lies in four points. First, forces of intimate somatic pairing appear rigorously localized in their expression to relatively minute terminal regions of the autosomes in mitosis (Figs. 1, 4). Second, these special regions provide the mechanism of diakinetic association of homologous autosomes (Fig. 5). Third, the diakinetic connections between autosomal homologues, like chiasmata, give origin to the metaphase associations of the first spermatocyte (Fig. 6). Fourth, the sex chromosomes have evolved a mechanism of segregation employing neither chiasmata nor somatic pairing (Figs. 3, 5, 6, 7). In this respect the sex chromosomes behave similarly to those of certain Nematocera.

Clearly, as in Darlington's view for *Drosophila* and other Brachycera, the pairing of the autosomes at meiosis in *Melophagus* must probably be considered to be due to exaggerated somatic pairing. It differs from the reported association of *Drosophila* in that in *Melophagus* the pairing is strictly localized terminally, and results in a pronounced, continuous bridge between homologous autosomes of a bivalent. Contrary to the case of *Drosophila*, repulsions of the kinetochores at meiosis in *Melophagus* take place much prior to metaphase and the paired connections between homologous autosomes become greatly drawn out. As the metaphase association of *Melophagus* follows one in which these obvious physical connections exist between the homologous autosomes, it is accordingly not unreasonable that, despite an optical void between the members of a bivalent, the association is still maintained by their sharing a portion of common pellicle or by connection of homologous halves to the same spindle elements. The anomaly of repulsion of homologues in exact equilibrium with persistent specific attraction among all four chromatids of a bivalent need not be invoked to account for the configuration of these bivalents of *Melophagus* at metaphase.[7]

The most significant features in the meiosis of *Melophagus* with regard to the general theory of meiosis in flies, however, lies in the following three points. (1) The homologous autosomes are only in contact over a short portion of their length, and this is assuredly not the whole of their homologous regions (Figs. 5, 6). (2) Parts of homologues not paired—in spite of their homology—lie apart at diakinesis (Fig. 5). (3) At prometaphase (Fig. 6) and metaphase there appears to be a state of tension between the spindle attachments and the points of association.[8] Briefly, the autosomal bivalents of *Melophagus* give evidence of the very characteristics of structure and behavior that Darlington holds to be indicative of chiasmata in the sex chromosomes of *Drosophila*. Nevertheless, in *Melophagus* they are probably all simulacra of chiasmata. As the relatively minute pairing regions of *Melophagus* seem to be merely chromosomal foci to which somatic pairing forces are restricted, and which by exaggeration of these forces hold bivalents together at meiosis, we are not as yet forced to assume with

Darlington a cytological need to suppose reciprocal chiasmata in the inert regions of the sex chromosomes of *Drosophila*.[9] Interstitial localization of somatic pairing forces at meiosis would give the figures of association that Darlington, Dobzhansky and others have published for the sex chromosomes of *Drosophila*, without involving chiasmata.

I wish to express my gratitude to Prof. Franz Schrader for suggesting this problem to me, to Prof. Schrader and Prof. Th. Dobzhansky for extended critical discussion of the results and to Mr. J. Monahan, foreman of the New York Stock Yards, for making it possible for me to easily obtain large numbers of *Melophagus*.

[1] Darlington (1931, see dia. III) first supposed the kinetochore to be included between the reciprocal chiasmata. This view can no longer be maintained, and it must now be supposed that the reciprocal chiasmata of the sex chromosomes are both on one side of the kinetochore.

[2] Darlington (1934) followed Imms, who relegated the Pupipara to a separate series of Brachycera.

[3] This is contrary to Lassman's report (1936) that the maturating egg and somatic cells show the chromosome number of *Melophagus ovinus* to be 16 ($n = 8$). Gonial plates of both male and female *Melophagus* show the diploid number to be 18. Apparently *Melophagus* shares with *Anthrax sinuosa* Wied. (Metz, 1916) the largest diploid number yet reported in the Diptera.

[4] In figures 1 and 4 the chromosome pairs at 12 o'clock and 11 o'clock, respectively, appear to be in lengthwise association. Actually the members of each pair are divergent.

[5] These connecting threads strongly recall those figured by Bauer (1931) for mitotic (gonial) prophases of *Tipula*.

[6] It must be asserted with Metz (1926), Huettner (1930) and Dobzhansky (1934) that it is not at all certain that the customary synaptic stages have been conclusively shown to be absent in Brachycera. Indeed, if Darlington is correct in his identification of chiasmata in the sex chromosomes of *Drosophila*, these chromosomes at least should reveal the customary sequence of stages.

[7] Presumably Darlington (1934, especially Fig. 36 of the side view of a diakinetic autosomal bivalent; 1937, especially pages 372, 498) has observed terminal connections or contacts between autosomal homologues at one extremity of the diakinetic bivalent of *Drosophila*. Whether these close associations, believed by Darlington to be due to pairing of the kinetochores, play a rôle similar to the autosomal pairing segments in *Melophagus* can only be conjectured, but it should be pointed out that Stevens (1908, Figs. 67, 69, 70, 71) and Dobzhansky (1934, Figs. 12, 18, 23, 35) have also published figures which call to mind the phenomena here described for *Melophagus*. It is interesting to note that Darlington (1931) used Stevens' figures in support of his earlier contention that fly autosomes form bivalents by reciprocal chiasmata. Each of these types of connection between autosomal bivalents deserves reinvestigation.

[8] The wording is nearly identical with that of Darlington (1934) describing the sex chromosome bivalents of *Drosophila*.

[9] It must be pointed out that Prof. Th. Dobzhansky believes that new evidence he has discovered in *Drosophila duncani* Sturt. supports Darlington's reciprocal chiasma hypothesis for the sex chromosomes of *Drosophila*.

Bauer, H., *Zeits. Zellforsch.*, **14**, 138 (1931).
Darlington, C. D., *Jour. Genet.*, **24**, 65 (1931).
Darlington, C. D., *Genetics*, **19**, 95 (1934).
Darlington, C. D., *Rec. Adv. Cytol.*, 2nd ed. (1937).
Dobzhansky, Th., *Zeits. Zellforsch.*, **21**, 169 (1934).
Frolova, S., *Ibid.*, **8**, 542 (1929).
Huettner, A. F., *Ibid.*, **11**, 615 (1930).
Keuneke, W., *Ibid.*, **1**, 357 (1924).
Lassman, G., *Ann. Ent. Soc. Am.*, **29**, 397 (1936).
Metz, C. W., *Jour. Exp. Zool.*, **21**, 213 (1916).
Metz, C. W., *Zeits. Zellforsch.*, **4**, 1 (1926).
Stevens, N. M., *Jour. Exp. Zool.*, **5**, 359 (1908).

SOME EFFECTS OF THE GENE W^v (DOMINANT SPOTTING) IN MICE

BY ELIZABETH FEKETE, C. C. LITTLE AND A. M. CLOUDMAN

ROSCOE B. JACKSON MEMORIAL LABORATORY

Communicated January 3, 1941

The gene *W* (dominant spotting) was identified by Little (1915) as having lethal effect when present in a homozygous condition. It was reported by de Aberle (1927) that among the offspring of matings in which both parents were heterozygous for the *W* gene, an inherited anemia occurred in certain animals. This appeared about the 16th day of intra-uterine life, became progressively more severe and resulted in the death of the animals within 10 days after birth. Little and Cloudman (1937) described a new dominant spotting mutation (*W'*) which probably represents a new allele of *W*. They stated that the homozygous offspring of this mutation may have a normal length of life, although they are usually completely or partially sterile. This new allele of *W* has been designated as W^v by The International Committee on Mouse Genetics Nomenclature (*Jour. Hered.*, in print). The *v* stands for viable and since it helps to express the nature of the mutant and leaves free for future use the less descriptive symbols *W'* and W^2, it will be used hereafter in this paper. Grüneberg (1939) conducted haematological studies on different alleles of *W*, including the new mutation W^v (*W'* of Little and Cloudman) which he designates W^2. He found that an inherited anemia existed and that the erythrocytes were reduced to about two-thirds of the normal values, while the white blood cells were reduced to a lesser extent. He established the normal values by doing blood counts on the normal litter mates of the W^vW^v animals.

The following is a report of the results of the microscopic examination of some of the organs of 3 male and 3 female mice which were homozygous for

the W^v gene. None of these animals had any litters although they were mated with their normal sibs. Emphasis, therefore, was placed on studies of the sex glands. The ovaries were serial sectioned and several sections were cut of the testes. Sections of the accessory male sex glands, and some of the endocrine glands, were also examined. Blood counts were taken on four of the animals.

Among the three males No. 7309 was 6 months, No. 5901, 21 months and No. 2655, 22 months old when killed. The testes of all of them were in normal position but were abnormally small. Microscopically the tunica albuginea of all of the glands was unusually thick. The seminiferous tubules of the two older animals were lined by Sertolian syncytium and seminiferous cells were not present. In the tubules of both testes of the 6 months old mouse, a few short areas showed spermatogenesis. Groups of spermatids and a few spermatozoa were present. In the glands of all three animals the interstitial cells were present in larger numbers than normal. The efferent ducts and the ductus epididymidis in the caput and corpus epididymidis were normal in structure, but contained no spermatozoa. The ductus epididymidis in the cauda was distended and contained a homogeneous secretion. The ductus deferens was normal. The vesicular glands were small, and contained a scant amount of secretion. The coagulating glands, the ventral and dorsal prostates, and the bulbo urethral glands did not show any remarkable deviation from the normal.

The thymus of ♂5901 was large and contained several unusually large cysts, partially lined by ciliated columnar epithelium. (Such cysts, although smaller in size, are frequently present in mice.) In the other two males this gland was normal in size and structure. The thyroids, parathyroids and adrenals were normal. The pituitaries of all three animals were slightly enlarged. Microscopically the anterior lobes showed hyperemia. Although cell counts were not made, the eosinophils (alpha cells) seemed to predominate and enlarged vacuolated basophils were not present.

Among the three females, No. 6772 was 11 months; the other two, Nos. 5588 and 5589, were 22 months old when killed. None of the ovaries of the females contained any normal ova. The ovaries of the two older animals were composed of dense stroma, granulosa cells and hypertrophied luteinlike cells. Many glandular areas lined by simple cuboidal (follicular) epithelium were present. In the left ovary of mouse No. 5588 some of these areas were cystic. The germinal epithelium showed numerous invaginations which gave the glands an irregular outline. The left ovary of mouse No. 6772 was unusually large and contained a cyst which was filled by an organized blood clot. A nodular area composed of densely arranged granulosa cells showing many mitotic figures as a sign of rapid growth was also present. In all three animals the epithelial cells of the fimbria and

ampulla of each oviduct showed secretory activity and many of the cells contained clear vacuoles. Grossly the uterine horns of all three animals appeared distended. Microscopically the mucosa of the endometrium showed a marked decrease in cellularity and the presence of a large amount of hyaline intercellular substance. The endometrial glands were considerably enlarged and some of them were cystic.

The adrenal glands of all three females showed some changes in the zona glomerulosa which were more pronounced in the older than in the younger females. Small groups of cells of this zone were hypertrophied; other small densely arranged cells projected in toward the cells of the zona fasciculata causing some lobulation, and disturbing the usual orderly arrangement of the cells of this zone. The thymus, thyroid and parathyroid glands did not show any abnormalities. The pituitaries were slightly enlarged and showed hyperemia.

The skin with the mammary glands in normal position was carefully removed from each of the three females. Observation with the aid of a dissecting microscope showed that the mammary glands consisted of fairly well-developed duct system possessing few side branches.

Blood counts were carried out on four animals. (The authors are indebted to Dr. L. W. Law and Dr. W. E. Heston for making the blood counts.) Table 1 gives the results.

TABLE 1

TOTAL ERYTHROCYTE AND LEUKOCYTE COUNT AND DIFFERENTIAL LEUKOCYTE COUNT OF W^vW^v MICE

NO. OF MOUSE	AGE IN MONTHS	ERYTHROCYTES PER MM.^{3}	LEUKOCYTES PER MM.^{3}	DIFFERENTIAL LEUKOCYTES IN %			
				LYMPHOCYTES	MONOCYTES	NEUTROPHILS	EOSINOPHILS
♂5901	21	4,848,000	3940	72	12	14	2
♀6772	11	3,264,000	8550	59	12	29	0
♀5588	22	3,200,000	3200	70	17	13	0
♀5589	22	3,264,000	Clotted	55	13	32	0
Average		3,644,000					

The average total erythrocyte count is lower in our animals than the figure given by Grüneberg. As he correctly states, the blood values vary strikingly with age; therefore, comparison of his age group (33–88 days with ours (11–22 months) is not advisable. To ascertain the approximate differences existing in normal younger and older animals, blood counts were taken on *C*57 black mice. The results show that a reduction of erythrocyte count can be expected in the older age groups. This reduction is comparable in the normal and in the abnormal (W^vW^v) animals. The results are summarized in table 2 and are self-explanatory.

TABLE 2

AVERAGE ERYTHROCYTE COUNT IN NORMAL AND W^vW^v MICE

STRAIN OF MICE	AGE	AVERAGE ERYTHROCYTE COUNT	NUMBER OF INDIVIDUALS
C57 black	77–86 days	9,038,000	13
C57 black	22–23 months	7,636,000	5
Grüneberg normals	33–88 days	9,030,000	8
Grüneberg W^vW^v	33–88 days	5,937,857	7
Little W^vW^v	11–21 months	3,644,000	4

Grüneberg stated that a reduction of erythrocytes to about one-half of the normal value characterizes the homozygous W^v animals. A similar reduction was found to be present in the W^vW^v animals reported on in this paper.

Little, C. C., *Am. Nat.*, **49**, 727–740 (1915).
de Aberle, S. B., *Am. Jour. Anat.*, **40**, 219–247 (1927).
Little, C. C., and A. M. Cloudman, *Proc. Nat. Acad. Sci.*, **23**, 535–537 (1937).
Grüneberg, Hans, *Genetics*, **24**, 777–810 (1939).

OPTICALLY RECIPROCAL GRATINGS AND THEIR APPLICATION TO SYNTHESES OF FOURIER SERIES

BY M. J. BUERGER

DEPARTMENT OF GEOLOGY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated January 2, 1941

1. Introduction.—Bragg[1] has recently called attention to the possibility of using the diffraction pattern from an appropriately constructed plane grating, to give a synthesis of the Fourier series commonly used to display the electron density distribution in a crystal. This use of diffraction is extremely limited due to the experimental lack of knowledge of the phases to assign to the individual holes in the grating, and the difficulty of experimentally controlling them when known. The writer[2] has pointed out that an extremely important field of usefulness for the diffraction synthesis of Fourier series exists in the syntheses of Patterson and Harker patterns, in both of which cases the phases are known. The application is particularly direct in case of the simple Patterson two-dimensional synthesis, in which case all the phases are the same and no special effort is necessary for their control.

Although the diffraction from simple plane gratings is well known, a discussion of gratings and their diffraction from a reciprocal lattice point of view has not been given.

2. Lattice Reciprocity of Plane Gratings.—An important characteristic of the diffraction from a plane grating is that the image of this diffraction is very approximately a lattice reciprocal to the grating producing it. This may be easily proven in the following way:

Let the original grating be designated G^*. (This starred designation is used because an important application of this diffraction is to produce a crystal lattice, G, from x-ray data in the form of its reciprocal lattice, G^*.) In figure 1, let a beam of parallelized monochromatic light of wave-length, λ, strike the grating at normal incidence. Consider, first, a single row of the grating (a horizontal row in G^*, Fig. 1). All holes in the row scatter in phase in a direction contained in a plane at right angles to the row. Now consider the conditions under which the several parallel rows will reënforce one another's diffraction within this range of freedom. In figure 2, it is evident that this occurs when the rows scatter in directions, δ, to the incident beam, such that

$$\sin \delta = \frac{n\lambda}{d^*}, \tag{1}$$

where d^* is the spacing of the rows and n is an integer. If these reënforced wavelets are focused by a lens, an image of bright spots occurs at the focal plane of the lens. The spots appear in the plane normal to the lattice rows in G^*, and have positions at distances, t, from the center, given by

$$t = D \tan \delta, \tag{2}$$

where D is the focal length of the lens. Combining (1) and (2) gives

$$t = D \tan \sin^{-1} \frac{n\lambda}{d^*}. \tag{3}$$

In actual practice, the spacing of the rows, d^*, is easily made of the order of a few millimeters, so that this is very large compared with λ. Consequently δ is very small and a very close approximation is

$$\sin \delta \approx \tan \delta. \tag{4}$$

Making this approximation, (3) becomes

$$t \approx (D\lambda)n \frac{1}{d^*}. \tag{5}$$

The terms in parenthesis are constant for the experiment, hence the several t's are integral multiples of the reciprocal of the row spacing, d^*. This means that the diffraction due to a set of rows in G^* is a lattice row in G normal to the original row. The translation interval, t, of the new row in G is the reciprocal of the spacing of the rows in G^*. The image row in G is

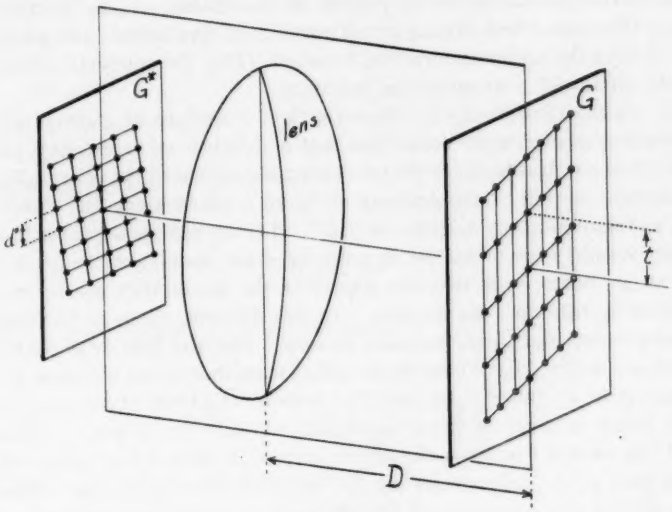


FIGURE 1

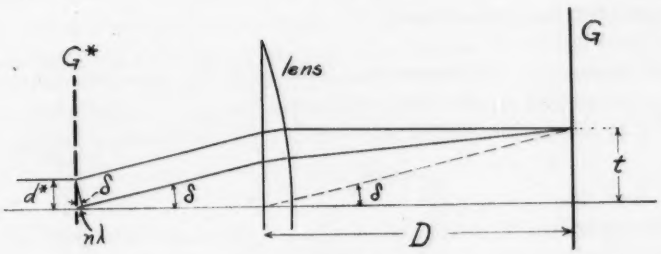


FIGURE 2

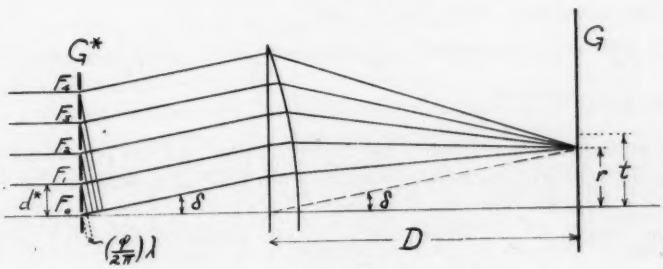


FIGURE 3

thus the reciprocal lattice equivalent of the spacing of the rows in G^*. Since this same effect occurs for all rows in G^* and follows the same law as (5) with the same experimental constant, $(D\lambda)$, the complete diffraction of the lattice G^* is its reciprocal lattice, G.

3. Pattern Reciprocity of Plane Gratings.—Reciprocal gratings are reciprocal in an even wider sense than that of dimensional reciprocity, for the pattern of one grating is the pattern reciprocal of that of the other. Figure 3 shows a section of the gratings of figure 1 taken along the direct ray line and normal to any desired row in G^*. Let us examine the contributions of the several rows of the set in some direction, making an angle, δ, with the direct beam, such that the phases of the waves they scatter do not necessarily reënforce one another. In this dirction, the zero row scatters a wave whose phase may be taken as zero. The first hole away scatters a wave in this direction whose phase differs from that of the zero hole by the phase angle, φ; this is equivalent to a fraction of a cycle equal to $\varphi/2\pi$ and to a linear measure of wave amounting to $(\varphi/2\pi)\lambda$. Figure 3 indicates that the second row from the origin scatters a wave whose phase differs with that of the origin wave by 2φ, the third hole by 3φ, etc. Further generalizing the conditions of the last section, let the several rows of the set have different transmissions: let the transmission of the central row be F_0, let the transmission of the first hole be F_1, etc. These features of the situation may be tabulated:

ROW NUMBER	ROW TRANSMISSION	PHASE ANGLE OF WAVE SCATTERED BY ROW	FRACTION OF CYCLE EQUIVALENT TO PHASE ANGLE
0	F_0	0	0
1	F_1	φ	$\frac{\varphi}{2\pi}$
2	F_2	2φ	$2\frac{\varphi}{2\pi}$
⋮	⋮	⋮	⋮
q	F_q	$q\varphi$	$q\frac{\varphi}{2\pi}$

The total contribution of all rows to the wave scattered in the direction, δ, is the summation of the Fourier series

$$S_\delta = F_0 e^{2\pi i(0)} + F_1 e^{2\pi i(\varphi/2\pi)} + F_2 e^{2\pi i(2(\varphi/2\pi))} + \ldots\ldots F_q e^{2\pi i(q(\varphi/2\pi))} \quad (6)$$
$$= \sum F_q e^{2\pi i(q(\varphi/2\pi))}.$$

This summation is performed by the diffraction, and the lens focuses the resulting wavelets originating from the rows in G^* at a point in G at a distance, r, from the direct beam record, and lying along a line at right angles to the rows in G^*. This distance, r, can be computed according to (2) as

$$r = D \tan \delta, \quad (7)$$

and the direction angle, δ, is given by a relation similar to (1), except that the integer, n, is replaced by a cycle fraction, $\varphi/2\pi$:

$$\sin \delta = \frac{\frac{\varphi}{2\pi}\lambda}{d^*}. \tag{8}$$

Combining (7) and (8) gives

$$r = D \tan \sin^{-1} \frac{\frac{\varphi}{2\pi}\lambda}{d^*}. \tag{9}$$

Making approximation (4), (10) becomes

$$r \approx D \frac{\frac{\varphi}{2\pi}\lambda}{d^*}. \tag{10}$$

Dividing (10) by (5), and letting $n = 1$, gives

$$\frac{r}{t} \approx \frac{\varphi}{2\pi}. \tag{11}$$

This means that the cycle fraction, $\varphi/2\pi$, may be replaced by a fraction, r/t, which represents the distance, r, as a fraction of the full translation, reciprocal to the row spacing in question. Making this replacement in (6), the summation becomes one pertaining to the point at a distance, r, in G, rather than a summation for the direction, δ, from the normal to G^*:

$$S_r \approx \sum F_q e^{2\pi i(q(r/t))}. \tag{12}$$

The diffraction due to the set of rows selected in G^* occurs not only at the single point having a central distance, r, in G, but at a continuum of points embracing all values of r, and forming a radial streak through the center of G. The full diffraction of the grating G^* comes not only from this particular row, but from all conceivable rows, no matter how complex. In this fashion, every radial streak in G is the pattern reciprocal of some conceivable set of rows in G^*.

4. *Pattern Synthesis in Terms of Lattice Coördinates.*—It is desirable to recast the results of the last section in terms of coördinates referred to lattice axes chosen in G^* and their reciprocals in G. If axes b^* and a^* are chosen in G^*, the diffraction due to sets of rows parallel with them produces radial streaks in G parallel to a and b, respectively. The summations for the particular point at a distance, x, along a and for the particular point at a distance, y, along b, are, according to (12),

$$S_x \approx \sum F_h e^{2\pi i(h(x/a))} \tag{13}$$

and

$$S_y \approx \sum F_k e^{2\pi i(k(y/b))} \tag{14}$$

We now require to know how each hole in G^*, as indexed on the axes, a^*, b^*, contributes to the summation at a point having absolute linear coördinates, x, y, with respect to the axes, a, b, in G. Figure 4 shows how a hole in G^*, having index hk, contributes to diffraction in some general direction. In this direction, the holes in neighboring positions along the a^* axis scatter with a difference in phase represented by the fraction of a cycle, $\varphi_{a^*}/2\pi$. The hth hole along the a^* axis therefore scatters a wave whose phase differs by $h(\varphi_{a^*}/2\pi)$ cycles from that of the origin hole. Similarly, the holes in neighboring positions along the b^* axis scatter with a difference in phase represented by the fraction of a cycle, $\varphi_{b^*}/2\pi$. The

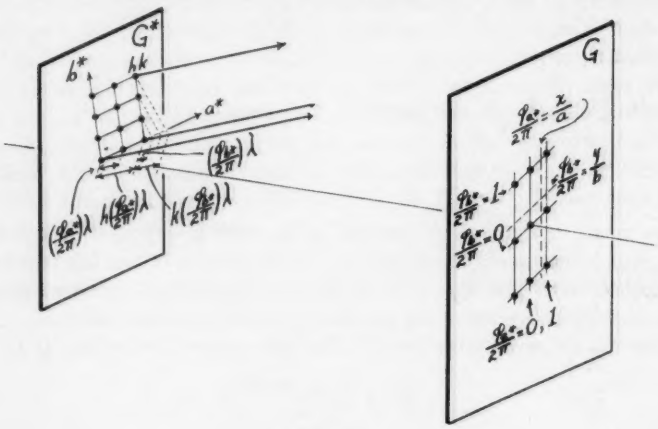


FIGURE 4

kth hole along the b^* axis therefore scatters a wave whose phase differs by $k(\varphi_{b^*}/2\pi)$ cycles from that of the origin hole. The total phase difference between the hole, hk, and the origin hole is therefore

$$P_{hk} = h\frac{\varphi_{a^*}}{2\pi} + k\frac{\varphi_{b^*}}{2\pi}. \tag{15}$$

From relations similar to (11),

$$\frac{x}{a} \approx \frac{\varphi_{a^*}}{2\pi}, \tag{16}$$

and

$$\frac{y}{b} \approx \frac{\varphi_{b^*}}{2\pi}. \tag{17}$$

Substituting these in (15), the phase of the hole, hk, with respect to that of the origin, may be very closely approximated by

$$P_{hk} \approx h\frac{x}{a} + k\frac{y}{b}. \tag{18}$$

Now this wave scattered by the hole, hk, is focused at the point having coördinates x, y in G. That this is the case can be easily grasped from the following reasoning: In (15), let $\varphi_{b^*}/2\pi$ successively take the particular constant values of 0, y/b, 1. For each of these values, let $\varphi_{a^*}/2\pi$ vary a little from zero. In this variation, the focused ray traces out a hyperbola on G because the locus of $\varphi_{b^*}/2\pi$ = constant is the surface of a cone which intersects G in a hyperbola. Since $(\varphi_{b^*}/2\pi)\lambda$ is exceedingly small compared with a^* in practical gratings, this hyperbola is substantially a straight line, a condition correlating with the approximation of (4). These several lines are labelled $\varphi_{b^*}/2\pi = 0$, $\varphi_{b^*}/2\pi = y/b$ and $\varphi_{b^*}/2\pi = 1$ in figure 4. In the same way, $\varphi_{a^*}/2\pi = 0$, $\varphi_{a^*}/2\pi = x/a$ and $\varphi_{a^*}/2\pi = 1$ may be mapped in figure 4. From this it follows that the wave having the phase indicated by (15) is focused at a point in G having relative coördinates x/a, y/b (i.e., at absolute coördinates, x, y). There is an inconsequential approximation involved here, but the approximation in no way affects the synthesis, but merely the position where the synthesis is focused.

From this discussion, it follows that at the point in the grating, G, having coördinates x, y with respect to the grating, axes, there occurs the synthesis,

$$S_{x,y} \approx \sum\sum F_{hk} e^{2\pi i\{h(x/a) + k(y/b)\}}. \tag{19}$$

This is in exactly the same form as the usual expression for the Fourier synthesis of the electron density at the point, x, y ,in the two-dimensional projection of a crystal cell. The corresponding Patterson synthesis can be obtained by making the holes in grating, G^*, proportional to $|F|^2$ rather than F, giving the summation

$$A_{x,y} \approx \sum\sum |F_{hk}|^2 e^{2\pi i\{h(x/a) + k(y/b)\}}, \tag{20}$$

which may also be written

$$A_{x,y} \approx \sum\sum |F_{hk}|^2 \cos^{2\pi\{h(x/a)+k(y/b)\}}. \tag{21}$$

5. *Control of Phases.*—In electron density syntheses, the transmission, F_{hk}, is, in general, a complex quantity usually of unknown phase. Two methods have been suggested[1, 2] for the control of this phase for the cases where the phase angle is known to be 0 or π. Any phase may be impressed on a term by placing in the path of each grating hole a thin sheet of plane-

parallel transparent material whose angle with respect to the transmitted beam may be varied, thus varying its effective optical thickness. Pieces cut from a single, perfect, mica cleavage are suggested for this purpose.

Another device useful in the control of phases is the utilization of multiple parallel gratings. Each gives a separate diffraction pattern, and these patterns may be optically combined, for example, with the aid of mirrors. The phase of an entire grating may be controlled, as in the last paragraph, with the aid of a single sheet. The use of two gratings having different phases may be applied to certain Harker syntheses, in which positive and negative terms must be summed at the same point of the pattern. The use of two (or more) gratings having different phases may also be applied to electron density summations, by permitting transmission of a given index, hk, to be contributed only from a single grating of the several used. For example, for centrosymmetrical projections, one grating may be made to have a positive phase, the other negative. A hole, hk, is uncovered on only one of the two gratings. This method may be extended to non-centrosymmetrical projections, by using several gratings; one grating transmits for the symmetrically equivalent h_1k_1's, another for h_2k_2's, etc. In this case, each grating has a general phase.

[1] Bragg, W. L., "A New Type of 'X-Ray Microscope,'" *Nature*, **143**, 678 (1939).

[2] Buerger, M. J., "The Photography of Interatomic Distances Vectors and of Crystal Patterns," *Proc. Nat. Acad. Sci.*, **25**, 383–388 (1939).

SEROLOGICAL REACTIONS WITH SIMPLE SUBSTANCES CONTAINING TWO OR MORE HAPTENIC GROUPS

BY LINUS PAULING, DAN H. CAMPBELL,* AND DAVID PRESSMAN

GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated January 13, 1941

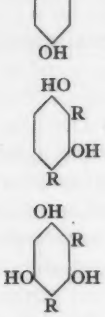
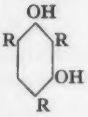
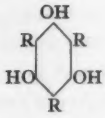
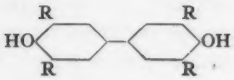
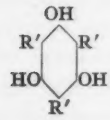
It has been inferred from inhibition studies that a simple hapten can combine with antibodies homologous to a conjugated protein involving that hapten[1] to give soluble compounds. In most cases haptens do not give the precipitin reaction with antiserum nor produce anaphylactic shock in animals sensitized with corresponding conjugated proteins. It was pointed out recently[2] that the framework theory of antibody-antigen precipitates suggests strongly that simple substances containing two or more haptenic groups should, under favorable circumstances, show all of the properties of more complex antigens; the thorough experimental test of this should provide evidence in support of or against the framework theory.

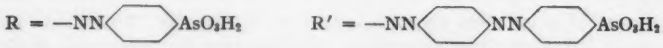
Landsteiner and van der Scheer[3] observed the occurrence of the precipitin reaction and anaphylaxis with simple substances formed by coupling two anilic acid molecules with resorcinol or tyrosine. On the other hand, Hooker and Boyd[4] have reported that they observed no formation of precipitate with several different compounds containing two haptenic groups. In order to obtain further evidence, we have prepared several substances with two or more haptenic groups per molecule and have tested them with suitable serum and with sensitized guinea pigs. We observed precipitate formation with each of the substances, and also observed anaphylaxis and desensitization with one of them, thus verifying the existence of the phenomena reported by Landsteiner and van der Scheer.

The serum used in our experiments was obtained from rabbits which had been given a course of six injections (alternately intraperitoneal and intravenous) of azoprotein per week for four weeks, the rabbits being bled nine days after the last injection. The azoprotein used was made by coupling sheep serum protein with diazotized arsanilic acid. Four azoprotein preparations, containing 0.68%, 1.9%, 2.3% and 3.3% arsenic (relative to total protein as determined from Kjeldahl nitrogen), respectively, were used. The titers of the serum of fourteen rabbits given one or another of these preparations were nearly the same, about 0.4 mg. antibody per ml., as found by Kjeldahl nitrogen analysis of the precipitate given with azoproteins made from egg albumin and arsanilic acid.

The haptens prepared and tested and the results obtained are shown in table 1.

TABLE 1

	SUBSTANCE	AMOUNT OF PRECIPITATE
I		None
II		Small
III		Medium
IV		Medium
V		Small
VI		Medium
VII		Large



The haptens I to VI were made by reaction of phenol, resorcinol, phloroglucinol, or *p,p'*-dihydroxybiphenyl with diazotized arsanilic acid, and VII by coupling with *p*-nitrosophenylarsonic acid the triamine obtained by the hydrolysis of the product of the reaction of phloroglucinol with diazotized *p*-aminoacetanilide.[5] The haptens were dissolved in alkali, brought to pH between 7 and 9, diluted and mixed with serum (0.5 ml. hapten solution to 0.4 ml. serum). At suitable concentrations of each of the haptens II to VII, between 10^{-4} and 10^{-6}, a noticeable precipitate could be seen in a few minutes, becoming heavier on standing overnight in the refrigerator. No

precipitate was obtained in control tests made with normal serum, and no precipitate was given by hapten I with antiserum. The optimum hapten concentration was in each case about 10^{-5} (0.01 mg. hapten per ml.).

Well-defined precipitates were obtained with substances II to VI, with no obvious dependence on variation of the number of haptenic groups from two to four. Substance VII gave a better reaction than the others; this we attribute to the smaller steric repulsion of antibody molecules attached by the long azobenzeneazophenylarsonate groups R′ than that of those attached by the shorter azophenylarsonate groups R.

The amount of antibody precipitated by substance VII under optimum conditions, about 0.4 mg. per ml. of serum, was nearly as great as that precipitated by azoprotein. We did not analyze the precipitates for hapten; with the assumption that all the hapten present is precipitated in the antibody-excess zone, our antibody analyses give for the antibody/antigen molecular ratio values between 0.2 and 0.9. These are smaller than the values found for protein antigens.

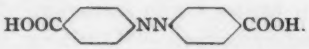
Guinea pigs actively sensitized by injection of azoprotein or passively sensitized by injection of rabbit anti-arsanilate serum were found to suffer shock on intravenous injection of $^1/_2$ ml. of solution containing 1 to 4 mg. of hapten VII, and to be desensitized with respect to azoprotein. Preliminary efforts to produce active antisera in rabbits and to sensitize guinea pigs by injection of hapten solutions gave negative results.

Our investigations confirm those of Landsteiner and van der Scheer. We attribute the antigenic properties of the substances studied to the presence in their molecules of two or more haptenic groups, and we consider that the phenomena provide strong support for the framework theory of serological precipitates, which provides a simple explanation of them.

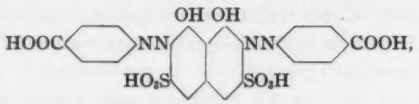
Landsteiner[6] has suggested as a possible alternative explanation that "the ready precipitability of these dyes is dependent upon peculiarities in constitution which, like those of fatty groups, diminish solubility in water and favor the formation of colloidal solutions." Landsteiner and van der Scheer reported somewhat better precipitations with hapten solutions which had aged by standing than with freshly prepared solutions. We, however, have observed no difference in behavior of old and fresh solutions. Our haptens dissolved easily at pH 9 to give clear stable solutions. At pH 7 the solutions precipitated on standing overnight, but mixtures with normal serum remained clear.

The failure of Hooker and Boyd to obtain precipitates with simple substances containing two haptenic groups may be attributed in part to the small size of the molecules of some of the substances studied by them, such as biphenyl-*p*,*p*′-dicarboxylic acid,

and *p,p'*-dicarboxylazobenzene,



Steric repulsion between two antibody molecules attached to such small molecules would be much stronger than for the molecules used by Landsteiner and van der Scheer and by us. This explanation does not apply to



for which they also report negative results.

Summary.—We have verified the observations of Landsteiner and van der Scheer that certain simple substances (azodyes) give precipitates with antisera and produce anaphylactic shock in sensitized animals. The substances found effective by these authors and by us contain two or more haptenic groups. The difference in properties of these substances and those containing only one haptenic group, which form only soluble compounds with antibodies, is explained in a simple way by the framework theory of serological precipitates, and the observations provide strong support for this theory.

* Fellow of the Rockefeller Foundation, 1940.

[1] Landsteiner, K., *The Specificity of Serological Reactions*, Charles C. Thomas, Baltimore (1936), Chap. V; *Biochem. Z.*, **93**, 117 (1919); **104**, 280 (1920); Marrack, J., and Smith, F. C., *Nature*, **128**, 1077 (1931); *British J. Exp. Path.*, **13**, 394 (1932); Haurowitz, F., and Breinl, F., *Z. physiol. Chem.*, **214**, 111 (1933).

[2] Pauling, L., *Jour. Am. Chem. Soc.*, **62**, 2643 (1940).

[3] Landsteiner, K., and van der Scheer, J., *Proc. Soc. Exp. Biol. and Med.*, **29**, 747 (1932); *Jour. Exp. Med.*, **56**, 399 (1932); **57**, 633 (1933); **67**, 79 (1938).

[4] Hooker, S. B., and Boyd, W. C., *Proc. Third Internat. Cong. Microbiol.*, p. 814 (1940); personal communication.

[5] We acknowledge the assistance of Mr. Miyoshi Ikawa and Mr. Carol Ikeda in the preparation of these compounds.

[6] Landsteiner, K., *The Specificity of Serological Reactions*, p. 120.

PROGRESSIVE IMMUNIZATION

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated January 3, 1941

In a recent note[1] we discussed the immunization of a population through contact of its individuals with a virus on the assumption that at the first contact the disease was acquired with a resulting lasting immunity. Such an all-or-none hypothesis with respect to the process of individual immunization may be sufficiently realistic for a disease like measles. Conceivably, however, a population might gain in immunity to some diseases through its individuals gaining progressively in immunity by successive contacts with the pathogenic agent—indeed as an extreme case it might be imagined that complete immunization was obtainable in that way without the development of cases of the disease.[2]

Suppose that if a person who has had no previous contact with the pathogen comes in contact with it, he has a chance p_1 of contracting the disease and acquiring a complete immunity and a chance $q_1 = 1 - p_1$ of not acquiring the disease but of gaining some immunity. When the person who has not had the disease comes a second time in contact with the pathogen he will have a chance p_2 of acquiring the disease with ensuing complete immunity and a chance q_2 of escaping but with the possibility of gaining somewhat further in immunity. The fractions of the persons just after the second contact will then be three:

p_1 who acquired complete immunity on first contact.
q_1p_2 who acquired complete immunity on second contact.
q_1q_2 who remain not completely immune.

Continuing the argument one finds four fractions after the third contact, namely, p_1, q_1p_2, $q_1q_2p_3$ who have had the disease and $q_1q_2q_3$ who have been thrice in contact without acquiring it but presumably have gained in immunity.

By the law of small numbers, if N contacts per capita have been spread over a population the fraction of persons who have had just k is $e^{-N}N^k/k!$. If there be dN further contacts per capita, the fraction of persons who have their $(k + 1)st$ contact is dN multiplied by the previous fraction, and of these the fraction $q_1q_2 \ldots q_k\, p_{k+1}$ will contract the disease and gain complete immunity. If $C(x)dx$ be the case rate during the time dx when dN contacts are distributed, we have[3]

$$C(x)dx = e^{-N} \sum_{k=0}^{k=\infty} q_1q_2 \ldots q_k p_{k+1} \frac{N^k}{k!} dN = \varphi(N)dN.$$

The average number N of contacts which have been experienced is itself a function of age x, but we need not make any special assumption with reference to it. The total number of cases per capita up to age x is

$$\psi(N) = \int_0^x C(x)dx = (1 - Q)(1 - e^{-N})$$
$$-e^{-N}\left[(q_1 - Q)N + (q_1q_2 - Q)\frac{N^2}{2!} + (q_1q_2q_3 - Q)\frac{N^3}{3!} + \ldots\right]$$

where Q stands for the indefinitely continued product[4] of the q's.

In no actual case do we know the values of p_i, but an assumption will be made merely in way of illustration.[5] Suppose

$$p_1 = 1/2,\ p_2 = 1/4, \ldots, p_k = 1/2^k$$
$$Q = \frac{1}{2}\cdot\frac{3}{4}\cdot\frac{7}{8}\cdots\frac{2^{k-1}}{2^k}\cdots \text{ converges to } Q = 0.2888.$$

Thus after an infinite number of contacts ($N = \infty$) the total fraction of the population which has acquired the disease is $1 - Q = 0.7112$ and the fraction $Q = 0.2888$ have acquired immunity progressively by repeated contact with the pathogen without ever contracting the disease. The values of $\varphi(N)$, $\psi(N)$ and the increment $\Delta\psi(N)$ for successive integral values of N from 0 to 7 are given in table 1. If the lapse of time (or age) be measured by the increase of the average number of contacts N, the column $\varphi(N)$ may be called the instantaneous case rate (per capita) by virtue of the relation $C(x) = \varphi(N)dN/dx$, the value of $\psi(N)$ may be called the total accumulated case rate, and the increments $\Delta\psi(N)$ are the case rates for the intervals from N to $N + 1$ measured by average contacts.

It may be recalled that on the all-or-none hypothesis, the corresponding formulas[1] are

$$\varphi(N) = e^{-N},\ \psi(N) = 1 - e^{-N},\ \Delta\psi(N) = e^{-N}(1 - 1/e).$$

Here $\log \varphi(N)$ and $\log \Delta\psi(N)$ plot as linear functions of N and $\psi(N)$ is a saturation curve which gives the fraction $I(N)$ of the population which has acquired immunity through having the disease. If we be dealing with the all-or-none hypothesis but with the complication, which is surely present in practice, that only a fraction of the cases are reported, whether this be current reporting of communicable diseases or results found by surveys which ask whether persons have had the disease,[6] and if we assume for simplicity that the under-reporting is independent of age (or of N) we should have to replace these formulas by

$$\varphi(N) = fe^{-N},\ \psi(N) = f(1 - e^{-N}),\ \Delta\psi(N) = fe^{-N}(1 - 1/e)$$

in terms of the supposed true condition in the population; such under-reporting could not fail to be present in respect to sub-clinical cases even if there were none with respect to the clinical cases. The only effect on our graphs would be that $\log_e \varphi$ would be $-N + \log f$ in place of $\log_e \varphi = -N$, and similarly for $\log \Delta\psi$; the lines would still be straight, but below their true positions when based on reported cases instead of on all cases.

If the values of $\log \varphi$ or of $\log \Delta\psi$ in table 1, as computed from our special hypothesis relative to progressive immunization of individuals, be

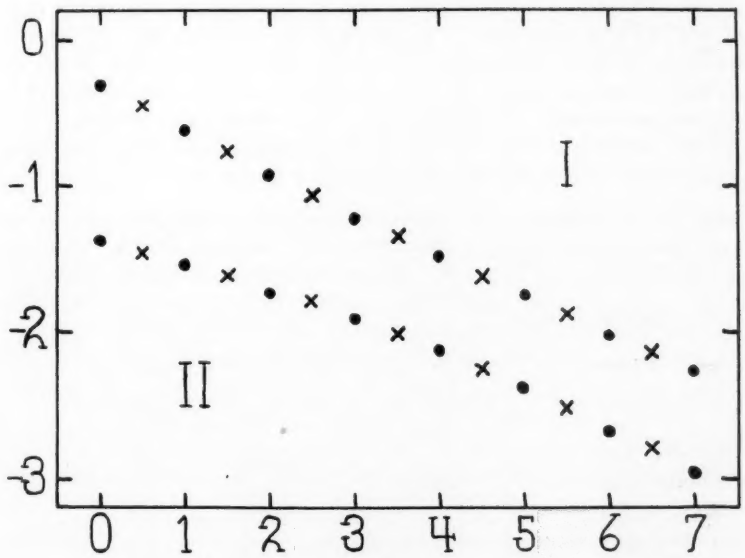


FIGURE 1

Graphs of $\log_{10} \varphi(N)$, solid dots, and $\log_{10} \Delta\psi(N)$, crosses, plotted against N. The upper graph I is for table 1 and the lower graph II is for table 2.

graphed, it will be observed that these lines are not far from straight (Fig. 1, I). Indeed with the inaccuracy of observed material what it is at present, even at best, the departures of the observations from any assumed law undoubtedly would be greater. Thus we may assert that if the population were immunized progressively as assumed, the observations could fairly be interpreted as due to the all-or-none hypothesis with fewer contacts and with under-reporting. Indeed the formulas

$$\varphi(N) = 0.4635e^{-0.6444\,N}, \quad \psi(N) = 0.7193(1 - e^{-0.6444\,N}),$$
$$\Delta\psi(N) = 0.7193e^{-0.6444\,N}(1 - e^{-0.6444})$$

fit the data in table 1 very well. If $N' = 0.6444\ N$, so that the contacts come only 0.6444 times as fast, and if the disease be considered only 71.93% reported, the results of the all-or-none hypothesis would practically duplicate[7] those in table 1.

Suppose in the second place that

$$p_1 = 0.04,\ p_2 = 0.03,\ p_3 = 0.02,\ p_4 = 0.01,\ p_i = 0,\ i > 4$$

so that there is never much chance of a case arising from contact with the pathogen, and no chance at all after four contacts[8] (because of the progressive acquirement of complete immunity). In table 2 are given the values of φ, ψ, $\Delta\psi$ and in figure 1, II, are graphed log φ and log $\Delta\psi$. The graphs are again not far from straight. The table could be explained as arising on an all-or-none hypothesis with $N' = 0.5165\ N$, i.e., with the contacts coming only about half as frequently, and with a reporting of only about ten per cent, whether because of failure to report clinical cases or because of the large number of sub-clinical cases which gave immunity.[9]

TABLE 1

N	φ	ψ	Δ
0	0.5000	0.000	
			0.353
1	0.2400	0.353	
			0.172
2	0.1190	0.525	
			0.086
3	0.0609	0.612	
			0.045
4	0.0321	0.657	
			0.024
5	0.0174	0.681	
			0.013
6	0.0097	0.694	
			0.007
7	0.0055	0.701	

TABLE 2

φ	ψ	$\Delta\psi$
0.0400	0.0000	
		0.0346
0.0293	0.0346	
		0.0245
0.0199	0.0590	
		0.0160
0.0125	0.0750	
		0.0098
0.0074	0.0848	
		0.0056
0.0041	0.0904	
		0.0030
0.0022	0.0934	
		0.0016
0.0011	0.0950	

For the case of this second example we give in table 3 against the average number of contacts, for the total population the number of persons per thousand who have had 0, 1, 2, 3 or 4 or more contacts and (in italics) the number who have had the disease. By our assumptions, those who have had four or more contacts are immune whereas those who have fewer contacts and have not had the disease are not yet completely immune.

These two specific hypothetical illustrations in which we have seen that progressive immunization simulates that of the all-or-none hypothesis with partial reporting are offered as a comment on our remark[1] that it is not necessary to infer from the similarity of age distribution of two diseases of childhood that the immunizing process relative to the individual is similar;

TABLE 3

NUMBER OF PERSONS PER THOUSAND WHO HAVE HAD A GIVEN NUMBER OF CONTACTS AND (IN ITALICS) NUMBER WHO HAVE HAD THE DISEASE

AVG. N	ACTUAL NUMBER OF CONTACTS 0	1	2	3	4+	TOTAL CASES	OTHER IMMUNES
1	*0*	*15*	*13*	*5*	*2*	*35*	*17*
	368	368	184	61	19	per thousand	
2	*0*	*11*	*19*	*16*	*14*	*59*	*129*
	135	271	271	180	143	per thousand	
3	*0*	*6*	*15*	*20*	*34*	*75*	*319*
	50	149	224	224	353	per thousand	
4	*0*	*3*	*10*	*17*	*55*	*85*	*512*
	18	73	147	195	567	per thousand	
5	*0*	*1*	*6*	*12*	*71*	*90*	*664*
	7	34	84	140	735	per thousand	
6	*0*	*1*	*3*	*8*	*82*	*93*	*767*
	2	15	45	89	849	per thousand	
7	*0*	*0*	*2*	*5*	*89*	*95*	*830*
	1	6	22	52	918	per thousand	

for one disease it may be all-or-none, for another it may be progressive with essentially similar curves of age distribution. Considerations other than these have to be brought to bear to make the decision—at least until the available material is decidedly better. As the all-or-none hypothesis is the simplest, one may appeal to the principle of Occam's razor to hold fast to it except as definite immunological observations require us to depart from it.[10]

[1] These PROCEEDINGS, **27**, 7–13 (1941). The assumption was made, and is continued in the present article, that the mortality due to the disease is negligible.

[2] For the purpose of this discussion a "case" corresponds to the completion of the acquisition of immunity all at once whether the case be clinical or sub-clinical and thus contrasts with a gain in immunity which remains only partial.

[3] Calculations will be made in terms of the average number of contacts and not in terms of age because of the added complication which arises from the uneven distribution of contacts by age discussed in our previous note.[1] The use of the law of small numbers implies a population homogeneous with respect to exposure except for the operation of chance.

[4] The product $Q = (1 - p_1)(1 - p_2) \ldots (1 - p_k) \ldots$ will converge to a value between 0 and 1 if the series $p_1 + p_2 \ldots + p_k + \ldots$ converges, but will converge to 0 if the series diverges. Thus if we had $p_1 = 1/2$, $p_2 = 1/3$, $\ldots$, $p_k = 1/k$, $\ldots$ we should have $Q = 0$ and everybody who lived long enough would contract the disease despite any immunity previously acquired by contact with the pathogen, the evidence of progressive immunity being merely the progressively lessened likelihood of getting the disease on successive contacts.

[5] The aim of this theory of progressive immunization with its illustration by specific examples is like that in Zinsser and Wilson, "Bacterial Dissociation and a Theory of the Rise and Fall of Epidemic Waves," *Jour. Preventive Med.*, **6**, 497–514 (1932), namely, to show the consequences of what might happen rather than to maintain that such a thing does happen.

[6] The factors of under-reporting may not be the same in the two cases; for example, surveys of persons for a history of measles generally give values of about 90% at ages 15

and over, which cannot be more than 10% under full reporting, whereas reports of cases to departments of public health are very much more deficient.

[7] If we assume $p_k = ar^{k-1}$, $a < 1$, $r < 1$, then

$$\varphi(N) = ae^{-N}\left[1 + q_1 rN + q_1q_2\frac{(rN)^2}{2!} + q_1q_2q_3\frac{(rN)^3}{3!} + \ldots\right] < ae^{-N}e^{rN}$$

since all q's are less than 1; but all products of q's are greater than Q, hence $\varphi(N) > aQe^{-N}e^{rN}$. Hence if θ be an appropriate fraction

$$\varphi(N) = ae^{-(1-r)N}[1 - \theta(1 - Q)].$$

Now if Q be near 1, i.e., if not many persons contract the disease (finish acquiring their immunity all at once), $\varphi(N)$ takes a form interpretable on the all-or-none hypothesis as due to contacts coming only $1/(1 - r)$ times as frequently with only the fraction a of cases recorded as such.

[8] Although four contacts completely immunize by hypothesis we do not consider all fourth contacts of persons who have not previously had the disease to be cases, but only 1% of them; in this there is a verbal inconsistency with the definition laid down in footnote 2.

[9] In neither of these illustrative examples is the relationship between log $\varphi(N)$ or log $\Delta\psi(N)$ and N strictly linear. It is possible to choose the successive values p_i in such a manner that the relationship is strictly linear, namely, $p_{i+1} = \alpha p_i/q_i$ with α any number less than q_1. In this case

$$\psi(N') = \frac{p_1}{1 - \alpha}(1 - e^{-N'}) \text{ with } N' = (1 - \alpha)N$$

and hence the case rates will be strictly as on the all-or-none hypothesis with contacts only $1 - \alpha$ times as frequent but with the cases reported only $p_1:(1 - \alpha)$ of the total which arise. For measles, if we assume 90% reporting (as on surveys) we have $p_1:(1 - \alpha) = 0.9$ and any values of p_1 and α which satisfy the linear equation $10p_1 = 9 - 9\alpha$ with $\alpha < 0.9$, will be satisfactory. These are such as

α	0	0.2	0.5	0.8	0.9
p_1	.9	0.72	0.45	0.18	0.090
p_2	0	0.51	0.41	0.18	0.089
p_3	0	0.21	0.35	0.17	0.088
p_4	0	0.05	0.26	0.16	0.087

and so on. The first column is "all-or-none" with 90% reporting; the others are "progressive" at various rates but also with 90% reporting.

If we had to rely on public health statistics where the reporting (though actually varying markedly with age) usually averages no more than 50% and if for illustrative purposes we use the average figure of 50% at all ages we have

α	0	0.2	0.5	0.8	0.9
p_1	.50	0.40	0.25	0.100	0.050
p_2	0	0.133	0.167	0.089	0.047
p_3	0	0.031	0.100	0.078	0.045
p_4	0	0.006	0.056	0.068	0.042

and so on. The results for $\alpha = 0.9$ imply frequent contacts (ten times as many as on the all-or-none hypothesis) but low and very slowly diminishing attack rates upon con-

tact, whereas those for $\alpha = 0.02$ imply only 25% more frequent contacts than the all-or-none hypothesis but with rather high though rapidly diminishing attack rates.

[10] Schuman and Doull, *Amer. Jour. Pub. Health*, **30**, Supplement to March, 1940, p. 21, state: "From these estimates of carrier prevalence and from the average annual increment in Schick-negatives, an estimate may be made of the number of carrier infections necessary, on the average, to produce sufficient antitoxin to give a negative Shick test." This seems to imply belief on their part that immunization as indicated by a negative Schick test may be acquired progressively; it is possible that a degree of immunity sufficient to prevent the disease arising on further contact is acquired before the Shick test turns negative, and if this degree of immunity were regularly acquired on the first carrier infection, we should classify the immunity as arising from a sub-clinical case under the definitions of footnote 2 and the phenomenon as to be considered of the all-or-none type.

ON DIRICHLET'S DIVISOR PROBLEM

BY AUREL WINTNER

DEPARTMENT OF MATHEMATICS, THE JOHNS HOPKINS UNIVERSITY

Communicated January 8, 1941

Let $d(n)$ denote the number of divisors of the positive integer n, and let

$$D(x) = d(1) + d(2) + \ldots + d([x]).$$

It is an elementary fact that this summatory function can approximately be represented by

$$F(x) = x(\log x + 2C - 1)$$

as $x \to \infty$, where C is the Euler-Mascheroni constant. The classical "divisor problem of Dirichlet" deals with the remainder term, $D(x) - F(x)$, of this elementary approximation. Dirichlet proved that

$$D(x) = f(x) + O(x^c), \quad \text{if } x \to \infty,$$

holds for $c = 1/2$; as indicated in his correspondence with Kronecker, he has known a sharper estimate also.[1] Today[2] it is known that the O-term is valid for a certain $c < 1/3$, and that it cannot be true for $c = 1/4$ (although it might hold for every $c > 1/4$). The present note leaves aside this difficult problem of estimation, and concerns itself with the fascinating fact that *the remainder term, $D(x) - F(x)$, admits of an anharmonic analysis* (on a proper scale).

Let $Z(s)$ be an abbreviation for the difference $Y_1(s) - H_1(s)$, where Y_1, H_1 denote the usual transcendents of the theory of cylinder functions. According to Voronoï,[3] the convergent explicit representation

$$\overline{D}(x) = F(x) + 1/4 + x^{1/2} \sum_{n=1}^{\infty} n^{-1/2} d(n) Z(4\pi n^{1/2} x^{1/2})$$

holds for every x, if $\overline{D}(x)$ is defined as the limit of $1/2\{D(x+h) + D(x-h)\}$ as h tends to zero. On applying to each of the terms $Z = Y_1 - H_1$ the asymptotic formulae of the cylinder functions Y_1, H_1, one easily obtains the convergent expansion

$$\overline{D}(x) = F(x) + 1/4 + \frac{x^{1/4}}{2^{1/2}\pi} \sum_{n=1}^{\infty} \frac{d(n)}{n^{3/4}} \cos(4\pi n^{1/2} x^{1/2} - \pi/4) + O(x^{-1/4});$$

(it is even known that the trigonometric series on the right is boundedly convergent on every finite x-range).

It seems to be natural to replace the independent variable x by $t = x^{1/2}$, and to ask whether or not the resulting *trigonometric series* in t is the *Fourier series* of the function to which it belongs in a formal fashion; that is, of the function

$$Q(t) = \{\overline{D}(t^2) - F(t^2)\}/t^{1/2}.$$

It turns out the answer is affirmative, in the sense that *the reduced remainder term, $Q(t)$, of Dirichlet's divisor problem is almost periodic* (B^2), $1 \leqq t < \infty$, *and has the Fourier series*

$$Q(t) \sim \frac{1}{2^{1/2}\pi} \sum_{n=1}^{\infty} \frac{d(n)}{n^{3/4}} \cos(4\pi n^{1/2} t - \pi/4).$$

This implies, in particular, that $Q(t)$ has an asymptotic distribution function.

It is clear from Bessel's identity that, if $M\{f(t)\}$ is an abbreviation for the average

$$\lim_{T \to \infty} \frac{1}{T} \int_1^T f(t)dt,$$

the italicized theorem is equivalent to the following pair of assertions:

(i) The amplitude belonging to an arbitrary frequency, that is, the Fourier average $M\{e^{i\lambda t} Q(t)\}$, exists for every real number λ, and is zero or has the value suggested by Voronoï's formal series according as λ is not or is of the form $\pm 4\pi n^{1/2}$, where $n = 1, 2, \ldots$.

(ii) The average intensity, $M\{Q(t)^2\}$, exists and is equal to the square sum of the amplitudes.

Cramér[4] deduced from the formal series a mean-value theorem which, in the above notations, may be expressed by saying that

$$\lim_{T \to \infty} \frac{1}{T} \int_1^T Q(t)^2 \left(\frac{t}{T}\right)^2 dt$$

exists and has a value compatible with (ii). Two partial integrations show that this result of Cramér is an Abelian consequence of (ii); in fact,

$$M\{Q(t)^2\} = \lim_{T \to \infty} \frac{1}{T} \int_1^T Q(t)^2 dt.$$

Thus the converse inference is Tauberian in nature. On the other hand, it seems to be hard to find a Tauberian theorem applicable to the present case. Consequently, a direct approach is needed.

Correspondingly, the proof of (i)-(ii), though straightforward, is somewhat lengthy, and will be published elsewhere. It applies also when Dirichlet's divisor problem, which can be thought of as the problem of lattice points in hyperbolas, is replaced by the problem of lattice points in circles.

[1] Lejeune-Dirichlet, P. G., *Werke*, **2**, 49–66 and 407 (1897).
[2] For references cf., e.g., *Enc. der Math. Wiss.*, II_3, 816–820.
[3] G. Voronoï, *Ann. Éc. Norm. Sup.*, ser. 3, **21**, 207–268 and 459–534 (1904).
[4] H. Cramér, *Math. Ztschr.*, **15**, 201–210 (1922).

UNIQUENESS THEOREMS FOR CONFORMAL MAPPING OF MULTIPLY CONNECTED DOMAINS

BY M. SHIFFMAN

DEPARTMENT OF MATHEMATICS, COLLEGE OF THE CITY OF NEW YORK

Communicated December 31, 1940

In a previous issue of these PROCEEDINGS,[1] it was proved on the basis of the Plateau problem that any k-fold connected domain is conformally equivalent to one of the following class of domains: The exterior of k non-intersecting closed curves which are, respectively, homothetic to k given convex analytic curves $C_1, C_2, \ldots, C_k$. It was furthermore possible to consider these domains normalized, e.g., for $k \geq 3$, by varying three of the boundaries C_1, C_2, C_3 only by radial transformations relative to a fixed interior point of each. We shall prove here that a k-fold connected domain can be mapped conformally onto a *unique* one of this class of normalized domains.

It suffices to prove the following: if $f(z)$ is an analytic function mapping one domain G of our class of normalized domains conformally on another normalized domain G', then $f(z) \equiv z$. The procedure is to suppose that the analytic function

$$g(z) = f(z) - z$$

is not identically zero (and consequently not identically constant), and to obtain a contradiction by counting the zeros and poles of $g(z)$.[2]

Let $C_1, C_2, \ldots, C_k$ be the boundaries of G, and $C_1', C_2', \ldots, C_k'$ of G'. Let N, P be the number of zeros and poles, respectively, of $g(z)$. It is easy to see that $P \leq 2$. For, either $f(z)$ takes ∞ into a finite point and finite point into ∞, in which case $P = 2$; or $f(\infty) = \infty$, for which $P = 1$ or 0. Concerning the zeros of $g(z)$ we have

$$N - P = \sum_{j=1}^{k} \frac{1}{2\pi i} \int_{C_j} d \log g(z) = \sum_{j=1}^{k} I_j$$

where the integration around C_j is taken in the clockwise direction, and where any zeros of $g(z)$ which occur on C_j are circumvented by protrusions outside C_j.

Now, $f(z)$ maps C_j into the curve C_j', obtained from C_j by a homothetic transformation; and as z moves clockwise once around C_j, the image point $f(z)$ moves clockwise once around C_j'. The complex number $g(z) = f(z) - z$ on C_j is the vector with initial point z on C_j and end-point $f(z)$ on C_j'. The vectors $g(z)$ form a vector field defined on the curve C_j. And if $g(z) \neq 0$ on C_j, I_j is exactly the *order* or *characteristic* of this vector field. We shall prove that $I_j \leq -1$ for $j = 1, 2, 3$, and $I_j \leq 0$ for $j > 3$.

Consider I_1. Since C_1' is obtained from C_1 by a radial transformation relative to an interior point O_1, and since C_1 is convex,[3] it follows that C_1' either surrounds C_1, or is interior to C_1, or coincides with C_1. In the first case, $I_1 = -1$. To prove this, contract the curve C_1 continuously till it approaches a point interior to C_1'. In this contraction, each vector remains attached to its position on C_1 and on C_1'. Since no vector ever becomes zero, the characteristic remains constant. In the limit the characteristic is -1, so that originally $I_1 = -1$.

A similar result applies if C_1' is interior to C_1. Finally, suppose that C_1' coincides with C_1. If $g(z)$ has no zeros on C_1, expand C_1'; $I_1 = -1$. If $g(z)$ has zeros on C_1, the circumvention of a zero will contribute an amount $-m/2$ to the integral I_1, where m is the order of the zero. Choose this circumvention to be an arc of a circle K of such small radius that the integral I_1 around it differs from $-m/2$ by less than ϵ, and such that $g(z)$ at each end-point of K differs from the tangential direction to C_1 by an angle less than ϵ. On an arc of C_1 between two successive zeros of $g(z)$, the change in angle differs by less than 2ϵ from the change in angle of the tangential rays on C_1. Thus, I_1 differs from $-1 - n/2$, where n is the sum of the orders of the zeros of $g(z)$ on C_1, by at most a fixed multiple of ϵ. Hence $I_1 = -1 - n/2$.

In all cases for the integral I_1, therefore, $I_1 \leq -1$. Similarly, $I_2 \leq -1$ and $I_3 \leq -1$.

Consider I_j for $j > 3$. There are three possible situations of C_j' relative to C_j (besides the possibilities enumerated above for C_1, which would yield $I_j \leq -1$):

1. C_j' and its interior lie completely outside C_j. The characteristic I_j of the vector field $g(z)$ is 0.

2. C_j' is tangent to C_j. If $g(z)$ has no zero on C_j, then C_j' can be displaced slightly so as not to intersect C_j. Hence $I_j = 0$ or -1 according as the tangency is external or internal. If $g(z)$ has a zero on C_j, a simple discussion of the circumvention of this zero shows that $I_j \leq -1$.

3. C_j' intersects C_j in two points, P and Q.[4] Denote the arcs of C_j interior and exterior to C_j' by α, β, respectively, and the arcs of C_j' interior and exterior to C_j by γ, δ. Let $f(z)$ map P, Q into the points P', Q' of C_j', and the arc α into α'. Suppose first that $g(z)$ has no zeros on C_j, so that P', Q' are distinct from P, Q, respectively. We shall distinguish three cases: either $\alpha' \subset \delta$ or $\alpha' \supset \delta$, or neither of these.

(a) $\alpha' \subset \delta$.—There is an arc $\bar{\alpha} \supset \alpha$ so that its image $\bar{\alpha}' \subset \delta$. By moving the points of C_j and of C_j' continuously so that $\bar{\alpha}$ approaches the whole curve C_j and $\bar{\alpha}'$ a point of C_j', the characteristic remains constant. In the limit, $I_j = 0$.

(b) $\alpha' \supset \delta$.—There is an arc $\bar{\bar{\alpha}} \subset \alpha$ so that its image $\bar{\bar{\alpha}}' \supset \delta$. Moving the points of C_j and of C_j' so that $\bar{\bar{\alpha}}$ approaches a point of C_j and $\bar{\bar{\alpha}}'$ the whole curve C_j', one sees in the limit that $I_j = -2$.

(c) Neither of (a), (b).—Move the arc α' continuously on C_j' so that its final position is interior to γ; this can be done, if necessary, by moving P' across Q or Q' across P. There is an arc $\bar{\alpha} \supset \alpha$ so that $\bar{\alpha}' \subset \gamma$. Letting $\bar{\alpha}$ approach the whole curve C_j and $\bar{\alpha}'$ a point of γ, we see that $I_j = -1$.

Finally, in (3), there remains the case when $g(z)$ has zeros on C_j. Suppose, e.g., that P' coincides with P. Then $I_j = -1$, -2 or -2 according as Q' is interior to δ, at Q, or interior to γ.

Therefore, in all cases for the integral I_j, where $j > 3$, $I_j \leq 0$.

Substituting the inequalities derived above in the formula $N = P + \sum_{j=1}^{k} I_j$, one obtains $N \leq 2 - 1 - 1 - 1 = -1$, contrary to $N \geq 0$. It follows that $f(z) \equiv z$, and the uniqueness theorem is proved.

Similarly, uniqueness can be established by the above method for other types of normalizations and other classes of normal domains.

[1] Courant, Manel and Shiffman, "A General Theorem on Conformal Mapping of Multiply Connected Domains," these PROCEEDINGS, 26, 503–507 (1940).

[2] After completing the present note, the author learned that this method was indicated by Carleman for the case of circular boundaries in a brief note in *Comptes rendus*, 168, 843–845 (1919).

[3] C_1 need only be star-shaped relative to the point 0_1.

[4] It can be shown that two homothetic convex analytic curves intersect in at most two points. If C_j were not convex, there may be more than two points of intersection, and the conclusion $I_j \leq 0$ would no longer be valid. For example, if C_j' intersected C_j in $2r$ points, then I_j may have values between $(-1-r)$ and $(-1+r)$, inclusive.